Ιόνιο Πανεπιστήμιο
Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών «Ψηφιακές Εφαρμογές & Καινοτομία»

Φοίβος Μυλωνάς
fmylonas@ionio.gr

«Τεχνολογίες Ευφυούς Διαχείρισης Ανθρωπιστικών Δεδομένων»
Μη-επιβλεπόμενη Μάθηση

Φοίβος Μυλωνάς, **fmylonas@ionio.gr**

Περιεχόμενα Α' μέρους

- Εισαγωγή
- Είδη μηχανικής μάθησης
- Μάθηση με επίβλεψη
- Μάθηση χωρίς επίβλεψη
- Κανόνες συσχέτισης
  - Αλγόριθμος A priori
- Clustering
  - Partion-based
  - Hierarchical
  - Probabilistic
- Αλγόριθμος DBSCAN

Εισαγωγή

- Η **μάθηση** σε ένα γνωστικό σύστημα, όπως γίνεται αντιληπτή στην καθημερινή ζωή, μπορεί να συνδεθεί με **2 βασικές ιδιότητες**:
  - την **ικανότητά στην απόκτηση γνώσης** κατά την αλληλεπίδρασή του με το περιβάλλον,
  - την **ικανότητά να βελτιώνει με την επανάληψη** τον **τρόπο εκτέλεσης** μία ενέργειας.
- Ορισμοί μάθησης:
  - Simon ('83), "η μάθηση σηματοδοτεί προσαρμοστικές αλλαγές σε ένα σύστημα με την έννοια ότι αυτές του επιτρέπουν να κάνει την ίδια εργασία, ή εργασίες της ίδιας κατηγορίας, πιο αποδοτικά και αποτελεσματικά την επόμενη φορά".
  - Minsky ('85), "... είναι να κάνουμε χρήσιμες αλλαγές στο μυαλό μας".
  - Michalski ('86), "... είναι η δημιουργία ή η αλλαγή της αναπαράστασης των εμπειριών".

Εισαγωγή

- Για τα συστήματα που ανήκουν στην συμβολική **Τεχνητή Νοημοσύνη** (ΤΝ), η μάθηση προσδιορίζεται ως **απόκτηση επιπλέον γνώσης**, που επιφέρει μεταβολές στην υπάρχουσα γνώση.

- Τα τεχνητά νευρωνικά δίκτυα (που ανήκουν στη μη συμβολική ΤΝ) έχουν δυνατότητα μάθησης μετασχηματίζοντας την εσωτερική τους δομή, παρά καταχωρώντας κατάλληλα αναπαριστάμενη γνώση.

Ορισμός Μηχανικής Μάθησης

- Ο άνθρωπος προσπαθεί να κατανοήσει το περιβάλλον του παρατηρώντας το και δημιουργώντας μια απλοποιημένη (αφαιρετική) εκδοχή του που ονομάζεται **μοντέλο** (**model**).
  - Η δημιουργία ενός τέτοιου μοντέλου, ονομάζεται **επαγωγική μάθηση** (**inductive learning**), ενώ η διαδικασία γενικότερα ονομάζεται **επαγωγή** (**induction**).
- Επιπλέον ο άνθρωπος έχει τη δυνατότητα να **οργανώνει** και να **συσχετίζει** τις **εμπειρίες** και τις παραστάσεις του δημιουργώντας νέες δομές που ονομάζονται **πρότυπα** (**patterns**).
- Η δημιουργία **μοντέλων** ή **προτύπων** από ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα, ονομάζεται **μηχανική μάθηση (machine learning)**.

Άλλοι ορισμοί

- Carbonell (1987), "... η μελέτη υπολογιστικών μεθόδων για την **απόκτηση νέας γνώσης**, **νέων δεξιοτήτων** και **νέων τρόπων οργάνωσης** της **υπάρχουσας γνώσης**".

- Mitchell (1997), "Ένα πρόγραμμα υπολογιστή θεωρείται ότι μαθαίνει από την **εμπειρία** Ε σε σχέση με μια κατηγορία **εργασιών** Τ και μια **μετρική απόδοσης** Ρ, αν η απόδοση του σε εργασίες της Τ, όπως μετριούνται από την Ρ, βελτιώνονται με την εμπειρία Ε".

- Witten & Frank (2000), "Κάτι μαθαίνει όταν **αλλάζει τη συμπεριφορά** του κατά τέτοιο τρόπο **ώστε να αποδίδει καλύτερα στο μέλλον**".

Είδη μηχανικής μάθησης

- Έχουν αναπτυχθεί πολλές τεχνικές μηχανικής μάθησης που χρησιμοποιούνται **ανάλογα με τη φύση του προβλήματος** και εμπίπτουν σε ένα από τα παρακάτω 2 είδη:
  - **μάθηση με επίβλεψη (supervised learning)** ή μάθηση με παραδείγματα (learning from examples),
  - **μάθηση χωρίς επίβλεψη (unsupervised learning)** ή μάθηση από παρατήρηση (learning from observation).

- Στη μάθηση **με επίβλεψη** το σύστημα καλείται να "μάθει" μια έννοια ή συνάρτηση από ένα σύνολο δεδομένων, η οποία αποτελεί περιγραφή ενός μοντέλου.

- Στη μάθηση **χωρίς επίβλεψη** το σύστημα πρέπει μόνο του να ανακαλύψει συσχετίσεις ή ομάδες σε ένα σύνολο δεδομένων, **δημιουργώντας πρότυπα**, χωρίς να είναι γνωστό αν υπάρχουν, πόσα και ποια είναι.

Μάθηση **με** Επίβλεψη

- Στη μάθηση με επίβλεψη το σύστημα πρέπει **να "μάθει" επαγωγικά** μια συνάρτηση που ονομάζεται **συνάρτηση στόχος** (**target function**) και αποτελεί έκφραση του μοντέλου που περιγράφει τα δεδομένα.
  - Η συνάρτηση στόχος χρησιμοποιείται για την πρόβλεψη της τιμής μιας μεταβλητής, που ονομάζεται **εξαρτημένη μεταβλητή** ή **μεταβλητή εξόδου**, βάσει των τιμών ενός συνόλου μεταβλητών, που ονομάζονται **ανεξάρτητες μεταβλητές** ή **μεταβλητές εισόδου** ή **χαρακτηριστικά**.

Μάθηση **με** Επίβλεψη

- Η επαγωγική μάθηση στηρίζεται στην "υπόθεση επαγωγικής μάθησης" (inductive learning hypothesis), σύμφωνα με την οποία:
  - Κάθε υπόθεση h που προσεγγίζει καλά τη συνάρτηση στόχο για ένα αρκετά μεγάλο σύνολο παραδειγμάτων, θα προσεγγίζει το ίδιο καλά τη συνάρτηση στόχο **και για περιπτώσεις που δεν έχει εξετάσει**.

- Διακρίνονται **2 είδη προβλημάτων** (learning tasks), τα **προβλήματα ταξινόμησης** και τα **προβλήματα παρεμβολής**.
  - Η **ταξινόμηση** (**classification**) αφορά στη δημιουργία μοντέλων πρόβλεψης **διακριτών τάξεων** (κλάσεων/κατηγοριών) (π.χ.: ομάδα αίματος).
  - Η **παρεμβολή** (**regression**) αφορά στη δημιουργία μοντέλων πρόβλεψης **αριθμητικών τιμών** (π.χ.: πρόβλεψη ισοτιμίας νομισμάτων ή τιμής μετοχής).

Μάθηση **χωρίς** Επίβλεψη

- Στη μάθηση **χωρίς** επίβλεψη το σύστημα έχει στόχο να **ανακαλύψει συσχετίσεις** και **ομάδες** από τα δεδομένα, **βασιζόμενο μόνο στις ιδιότητές τους**.
- Σαν αποτέλεσμα **προκύπτουν πρότυπα** (περιγραφές), κάθε ένα από τα οποία περιγράφει ένα μέρος από τα δεδομένα.
- Παραδείγματα προτύπων πληροφόρησης είναι οι κανόνες συσχέτισης (**association rules**) και οι συστάδες (**clusters**), οι οποίες προκύπτουν από τη διαδικασία της συσταδοποίησης (clustering).

1. Κανόνες Συσχέτισης

- Η **ανακάλυψη ή εξόρυξη κανόνων συσχέτισης** (association rule mining) εμφανίστηκε αρκετά αργότερα από τη μηχανική μάθηση και έχει περισσότερες επιρροές από την ερευνητική περιοχή των **βάσεων δεδομένων**.
- Προτάθηκε στις αρχές της δεκαετίας του '90 από τον Rakesh Agrawal ως τεχνική ανάλυσης καλαθιού αγορών (**market basket analysis**) όπου το ζητούμενο είναι η ανακάλυψη συσχετίσεων ανάμεσα στα αντικείμενα μιας βάσης δεδομένων.
- Στο συγκεκριμένο πρόβλημα υπάρχει ένας μεγάλος αριθμός αντικειμένων (items), π.χ. *ψωμί*, *γάλα*, κ.λ.π.. Οι πελάτες γεμίζουν τα καλάθια τους με κάποιο υποσύνολο αυτών των αντικειμένων και το ζητούμενο είναι να βρεθεί ποια από αυτά τα αντικείμενα αγοράζονται μαζί, χωρίς να ενδιαφέρει ποιος είναι ο αγοραστής.

Κανόνες Συσχέτισης

- Οι κανόνες συσχέτισης είναι προτάσεις της μορφής $\{X_1,...,X_n\} \rightarrow Y$, που σημαίνει ότι αν βρεθούν όλα τα X_1, ..., X_n στο καλάθι (στην ανάλυση καλαθιού αγορών) τότε είναι πιθανό να βρεθεί και το Υ.
- Για παράδειγμα, ένας τέτοιος κανόνας θα μπορούσε να λέει:
  - «όποιος αγοράζει καφέ (X_1) και ζάχαρη (X_2) αγοράζει και αναψυκτικά (Υ)»

Κανόνες Συσχέτισης

- Απλή αναφορά ενός τέτοιου κανόνα δεν έχει μεγάλη αξία αν δε συνοδεύεται από **κάποια ποσοτικά μεγέθη** που μετρούν την ποιότητα των ευρεθέντων κανόνων συσχέτισης.
- Τέτοια μεγέθη είναι η **Υποστήριξη** και **Εμπιστοσύνη**.

- **Υποστήριξη** (support) ή **κάλυψη** (coverage)
  - εκφράζει την πιθανότητα να βρεθεί το καλάθι $\{X_1, \ldots, X_n, Y\}$ στη βάση δεδομένων
  - ισούται με το λόγο των εγγραφών που περιλαμβάνουν το $\{X_1,\ldots,X_n, Y\}$ / το σύνολο των εγγραφών

Κανόνες Συσχέτισης

- Απλή αναφορά ενός τέτοιου κανόνα δεν έχει μεγάλη αξία αν δε συνοδεύεται από **κάποια ποσοτικά μεγέθη** που μετρούν την ποιότητα των ευρεθέντων κανόνων συσχέτισης.
- Τέτοια μεγέθη είναι η **Υποστήριξη** και **Εμπιστοσύνη**.

- **Εμπιστοσύνη** (confidence) ή **ακρίβεια** (accuracy)
  - εκφράζει την πιθανότητα να βρεθεί το Y σε ένα καλάθι που περιέχει τα $\{X_1, ..., X_n\}$
  - ισούται με το λόγο των εγγραφών που περιλαμβάνουν το $\{X_1, ..., X_n, Y\}$ / το σύνολο των εγγραφών που περιλαμβάνουν τα X_i.

Αλγόριθμοι Εύρεσης Κανόνων Συσχέτισης

- Για την ανακάλυψη κανόνων συσχέτισης χρησιμοποιείται η **ιδιότητα της μονοτονίας** (monotonicity property) ή αλλιώς **ιδιότητα a priori** σύμφωνα με την οποία:

"Αν ένα σύνολο αντικειμένων S είναι συχνό, τότε όλα τα υποσύνολα του S είναι επίσης συχνά".

- Π.χ. αν είναι συχνό το {γάλα, ψωμί, λάδι}, τότε είναι τουλάχιστον εξίσου συχνό και το {γάλα, ψωμί} (ή το {γάλα, λάδι}, ή το {ψωμί, λάδι}).

Αλγόριθμοι Εύρεσης Κανόνων Συσχέτισης

- **Συχνό** είναι ένα σύνολο αντικειμένων όταν εμφανίζεται σε ποσοστό των καλαθιών **ίσο ή μεγαλύτερο από ένα όριο** που συνήθως ορίζει ο χρήστης.

- Σε έναν αλγόριθμο εύρεσης κανόνων συσχέτισης μας ενδιαφέρει κυρίως **ο αριθμός των περασμάτων** στα δεδομένα που απαιτείται κατά την εκτέλεσή του.

Αλγόριθμος «A priori»

- Προτάθηκε από τον **Rakesh Agrawal** το **1994** και είναι ίσως ο **κλασικότερος αλγόριθμος ανακάλυψης κανόνων συσχέτισης**.

- Περιλαμβάνει **2** βασικά **βήματα**
  - **δημιουργία των συχνών συνόλων αντικειμένων**
  - **δημιουργία των κανόνων συσχέτισης**.

Αλγόριθμος «A priori»

- Η **διαδικασία της δημιουργίας συχνών συνόλων** αντικειμένων περιλαμβάνει 2 στάδια:
  - **Αρχικά** δημιουργείται ένα **σύνολο υποψήφιων συχνών αντικειμένων C_i**
  - Στην συνέχεια χρησιμοποιώντας το **όριο υποστήριξης** (**support**), δημιουργείται το **σύνολο των συχνών συνόλων αντικειμένων L_i**.
  - Η διαδικασία **επαναλαμβάνεται** πραγματοποιώντας **διαδοχικά περάσματα** στα δεδομένα μέχρι να βρεθούν
    - είτε τα **συχνά σύνολα αντικειμένων** ενός **προκαθορισμένου** επιπέδου ή
    - τα **μέγιστα συχνά σύνολα αντικειμένων**.
  - Το πρώτο αυτό στάδιο επιπλέον αποτελείται από ένα **βήμα συνένωσης** (**join step**) και ένα **βήμα κλαδέματος** (**prune step**), τα οποία συνήθως εκτελούνται στη μνήμη και έτσι δεν είναι ιδιαίτερα χρονοβόρα.

Αλγόριθμος «A priori»

- Στο 2ο στάδιο, για τη **δημιουργία των κανόνων συσχέτισης** ελέγχεται η **εμπιστοσύνη** (**confidence**) όλων των πιθανών κανόνων που προκύπτουν από τα μέγιστα συχνά σύνολα αντικειμένων

- Στο τέλος μένουν εκείνοι οι κανόνες των οποίων η εμπιστοσύνη **ξεπερνά το όριο που τέθηκε** από το χρήστη.

Δημιουργία **Συχνών Συνόλων Αντικειμένων**

- Έστω ότι το όριο υποστήριξης είναι sup. Στο **1ο πέρασμα** βρίσκονται τα αντικείμενα εκείνα που εμφανίζονται στη βάση δεδομένων **σε ένα ποσοστό sup** των καλαθιών **ή μεγαλύτερο**.
  - Αυτό το σύνολο ονομάζεται **σύνολο συχνών αντικειμένων** (frequent 1- itemset) και συμβολίζεται με L1.
  - Από το L1 προκύπτουν, κάνοντας **όλους τους δυνατούς συνδυασμούς**, τα υποψήφια (συχνά) **ζεύγη αντικειμένων C2**.
- Στο **2ο πέρασμα**, τα ζεύγη του C2 των οποίων το πλήθος **ικανοποιεί** το κριτήριο **sup**, δημιουργούν το L2, δηλαδή τα συχνά ζεύγη (frequent 2-itemsets).
  - Από το L2 προκύπτουν οι υποψήφιες (συχνές) **τριάδες αντικειμένων C3** που θα χρησιμοποιηθούν στο τρίτο πέρασμα.
  - Οι υποψήφιες τριάδες C3 είναι σύνολα του τύπου {A,B,C} τέτοια, ώστε **όλα τα υποσύνολά του μεγέθους δύο, δηλαδή τα {A,B}, {A,C}, {B,C}, να ανήκουν στο L2.**
- Στο **3ο πέρασμα**, υπολογίζεται ο αριθμός των εμφανίσεων των τριάδων του C3 και **με βάση το κριτήριο sup** δημιουργείται το **L3**.
- Η διαδικασία συνεχίζεται για προκαθορισμένο αριθμό επιπέδων ή μέχρι να αδειάσουν τα υποψήφια συχνά σύνολα αντικειμένων και να δημιουργηθούν τα μέγιστα συχνά σύνολα αντικειμένων.

Παράδειγμα Εφαρμογής του Apriori

- Έστω ένα σύνολο δεδομένων που αντιστοιχούν σε 10 διαφορετικά καλάθια αγορών από ένα super market.
- Κάθε καλάθι περιλαμβάνει ένα υποσύνολο των προϊόντων του super market.
- Για παράδειγμα, στο **1ο καλάθι** ο πελάτης αγόρασε μόνο **ψωμί** και **γάλα**.

Καλάθι	Ψωμί	Καφές	Γάλα	Ζάχαρη
#1	1	0	1	0
#2	0	1	0	0
#3	1	0	1	1
#4	0	1	0	1
#5	1	0	1	1
#6	1	1	1	0
#7	1	0	0	1
#8	1	1	1	1
#9	0	0	1	1
#10	1	1	0	1

Παράδειγμα Εφαρμογής του Apriori – 1ο σκέλος

- Έστω επίσης ότι η ζητούμενη **υποστήριξη** είναι **sup=40%** και η ζητούμενη **εμπιστοσύνη conf=80%**.
- Στο 1ο πρώτο βήμα, ο Apriori **υπολογίζει την υποστήριξη όλων των αντικειμένων**, δηλαδή δημιουργεί το σύνολο L1.
  - S{Ψωμί} = 7/10 = 70% ≥ sup
  - S{Καφές} = 5/10 = 50% ≥ sup
  - S{Γάλα} = 6/10 = 60% ≥ sup
  - S{Ζάχαρη} = 7/10 = 70% ≥ sup
- Συνεπώς L1={ Ψωμί, Καφές, Γάλα, Ζάχαρη }
- Στο 2ο βήμα, παράγονται **όλοι οι δυνατοί συνδυασμοί** των αντικειμένων, για να δημιουργηθεί το **σύνολο υποψήφιων ζευγών αντικειμένων**, δηλαδή το σύνολο C2.
  - C2={ {Ψωμί,Καφές},{Ψωμί,Γάλα},{Ψωμί,Ζάχαρη},{Καφές,Γάλα},{Καφές,Ζάχαρη},{Γάλα,Ζάχαρη} }

Παράδειγμα Εφαρμογής του Apriori – 1ο σκέλος

- Κατόπιν, **υπολογίζεται η υποστήριξη** των μελών **του C2** και απορρίπτονται εκείνα που δεν ξεπερνούν το όριο ελάχιστης υποστήριξης, ώστε να **δημιουργηθεί το σύνολο συχνών ζευγών L2**.
  - ({Ψωμί, Καφές}) = 3/10 = 30% < sup (απορρίπτεται)
  - S({Ψωμί, Γάλα}) = 5/10 = 50% ≥ sup
  - κτλ.
  - …

- Τελικά: L2={{Ψωμί,Γάλα},{Ψωμί,Ζάχαρη},{Γάλα,Ζάχαρη}}

Παράδειγμα Εφαρμογής του Apriori – 1ο σκέλος

- Από το L2, με τον ίδιο τρόπο, θα δημιουργηθούν τα C3 και L3 (**αν τελικά υπάρχουν "συχνές" τριάδες**). Στο συγκεκριμένο παράδειγμά, το βήμα δημιουργίας υποψήφιων τριάδων έχει ως εξής:
- Βήμα **συνένωσης**: {Ψωμί,Γάλα}U{Ψωμί,Ζάχαρη}= {Ψωμί,Γάλα,Ζάχαρη}
- Βήμα **κλαδέματος**: Οι επιμέρους δυάδες (ζεύγη) του {Ψωμί,Γάλα,Ζάχαρη} ανήκουν όλες στο L2, άρα:
  - C3={{Ψωμί,Γάλα,Ζάχαρη}}
  - S({Ψωμί,Γάλα,Ζάχαρη})=3/10=30% (απορρίπτεται), άρα L3={}.

- Ο αλγόριθμος εύρεσης συχνών συνόλων αντικειμένων **σταματά εδώ** και συνεπώς το **μέγιστο συχνό σύνολο αντικειμένων είναι το L2**.

Παράδειγμα Εφαρμογής του Apriori – 2ο σκέλος

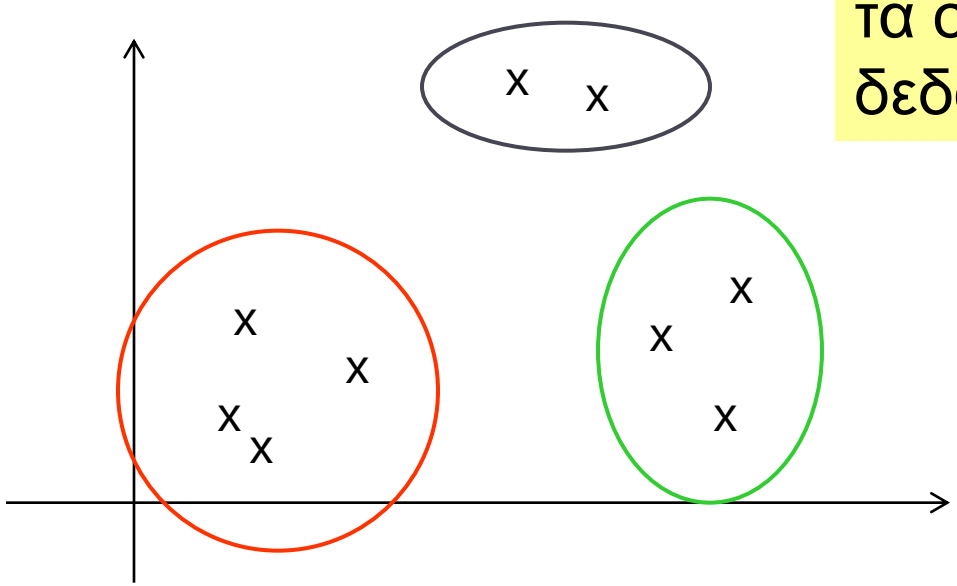
- Το επόμενο βήμα είναι η εξαγωγή των κανόνων από τα συχνά σύνολα (**στο συγκεκριμένο παράδειγμα μόνο το L2**), βάσει της εμπιστοσύνης τους.
  - L2={{Ψωμί,Γάλα},{Ψωμί,Ζάχαρη},{Γάλα,Ζάχαρη}}
- **Ελέγχεται η εμπιστοσύνη** όλων των πιθανών κανόνων που μπορεί να προκύψουν από το L2.
- {Ψωμί,Γάλα}
  - Ψωμί→Γάλα: εμπιστοσύνη = 5/7 = 71% < conf (απορρίπτεται)
  - Γάλα→Ψωμί: εμπιστοσύνη = 5/6 = 83% > conf (**εγκρίνεται**)
- {Ψωμί,Ζάχαρη}
  - Ψωμί→Ζάχαρη: εμπιστοσύνη = 5/7 = 71% < conf (απορρίπτεται)
  - Ζάχαρη→Ψωμί: εμπιστοσύνη = 5/7 = 71% < conf (απορρίπτεται)
- {Γάλα,Ζάχαρη}
  - Γάλα→Ζάχαρη: εμπιστοσύνη = 4/6 = 66% < conf (απορρίπτεται)
  - Ζάχαρη→Γάλα: εμπιστοσύνη = 4/7 = 57% < conf (απορρίπτεται)
- Τελικά παράγεται μόνο ο κανόνας: Γάλα→Ψωμί, δηλαδή **όποιος αγοράζει Γάλα αγοράζει και Ψωμί.**
  - Σημ.: Αν ελαττώσουμε τη ζητούμενη εμπιστοσύνη στο 70%, τότε θα παραχθούν 4 κανόνες.

2. Clusters

- Είναι πρότυπα πληροφόρησης που προκύπτουν με συσταδοποίηση (clustering) δηλαδή διαχωρισμό ενός συνόλου (συνήθως πολυδιάστατων) δεδομένων σε συστάδες, ώστε:
  - **σημεία** που ανήκουν στην **ίδια συστάδα** να **μοιάζουν όσο το δυνατόν περισσότερο** και
  - **σημεία** που ανήκουν σε **διαφορετικές συστάδες** να **διαφέρουν όσο το δυνατόν περισσότερο**.
- Χωρίς επίβλεψη: δεν παρέχεται καμία πληροφορία στον αλγόριθμο σχετικά με το ποια σημεία δεδομένων ανήκουν σε ποιες συστάδες.

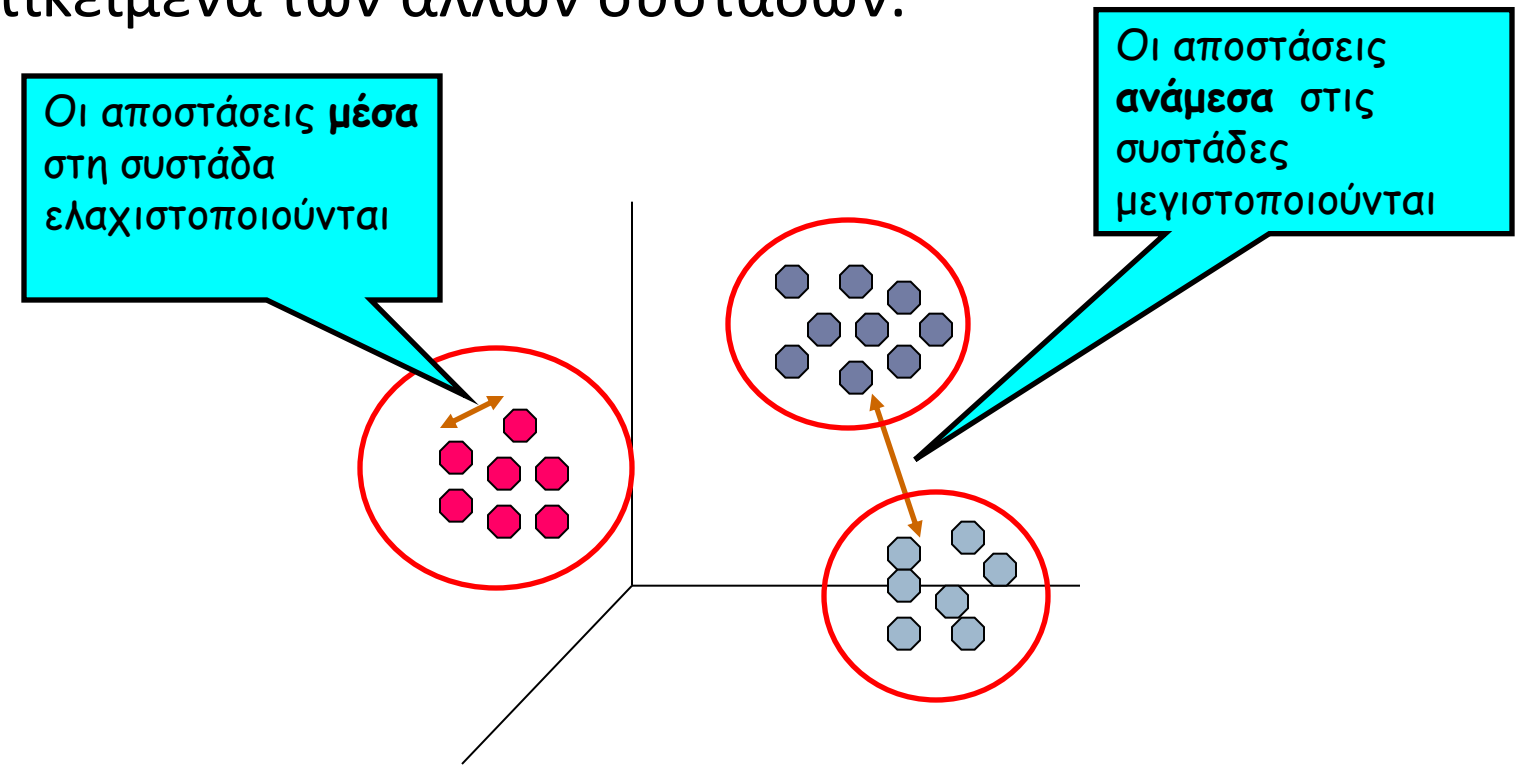
Clusters

Ποιες θα είναι οι συστάδες για αυτά τα σημεία δεδομένων?



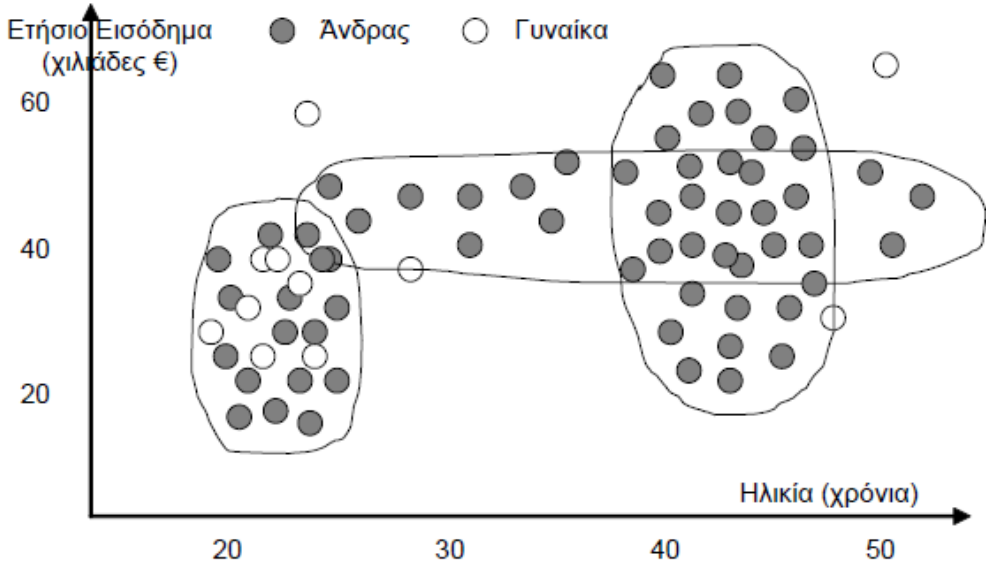
Τι είναι συσταδοποίηση

- Εύρεση συστάδων αντικειμένων έτσι ώστε τα αντικείμενα σε κάθε συστάδα να είναι όμοια (ή να σχετίζονται) και διαφορετικά (ή μη σχετιζόμενα) από τα αντικείμενα των άλλων συστάδων.



Clusters

- Στο Σχήμα απεικονίζεται γραφικά μία υποθετική συσταδοποίηση σε δεδομένα αγοραστών αυτοκινήτων, με βάση την ηλικία (άξονας x), το ετήσιο εισόδημα (άξονας y) και το φύλο. Διακρίνονται τρεις ομάδες:
  - "αγοραστές νεαρής ηλικίας ανεξαρτήτως φύλλου",
  - "άνδρες αγοραστές με υψηλό εισόδημα, όλων των ηλικιών μέχρι τα 53 χρόνια" και
  - "άνδρες αγοραστές ηλικίας περίπου 44 ανεξαρτήτως εισοδήματος".



Αλγόριθμοι Συσταδοποίησης

- Υπάρχουν 2 + 1 = 3 γενικές κατηγορίες αλγορίθμων συσταδοποίησης:

- Οι αλγόριθμοι βασισμένοι σε διαχωρισμούς (**partition based**), που προσπαθούν να βρουν τον **καλύτερο διαχωρισμό ενός συνόλου δεδομένων** σε ένα **συγκεκριμένο** αριθμό συστάδων.
- Οι ιεραρχικοί (**hierarchical**) αλγόριθμοι, που προσπαθούν **με ιεραρχικό τρόπο** να ανακαλύψουν τον **αριθμό** και τη **δομή** των **συστάδων**.
- Οι πιθανοκρατικοί (**probabilistic**) αλγόριθμοι, που βασίζονται σε **μοντέλα πιθανοτήτων**.

Συσταδοποίηση

- Η συσταδοποίηση απαιτεί κάποιο **μέτρο της ομοιότητας ή διαφοράς μεταξύ των δεδομένων**. Συνήθως υπολογίζεται η "απόσταση" μεταξύ των δεδομένων.
- Έστω ένα σύνολο δεδομένων D, και δύο δεδομένα του, x, y που περιγράφονται από m το πλήθος χαρακτηριστικά: $(x_1,x_2,\ldots,x_m)$, $(y_1,y_2,\ldots,y_m)$.
- Τυπικά μέτρα απόστασης αυτών των 2 δεδομένων είναι η απόσταση Μανχάταν ή η Ευκλείδεια απόσταση.

$$d(x,y)=\sum_i |x_i - y_i|$$

Απόσταση *Μανχάταν*

$$d(x,y)=\sqrt{\sum_i (x_i - y_i)^2}$$

Ευκλείδεια απόσταση

Συσταδοποίηση

- Αν κάποια χαρακτηριστικά είναι διακριτά, τότε η απόσταση των τιμών τους θεωρείται:
  - 0, αν πρόκειται για την **ίδια τιμή**
  - 1, αν πρόκειται για **διαφορετικές τιμές**

- Τα αριθμητικά χαρακτηριστικά θα πρέπει να ομογενοποιούνται ώστε η απόστασή τους να πέφτει μέσα στο διάστημα [0,1].

Α. Αλγόριθμοι Βασισμένοι σε Διαχωρισμούς

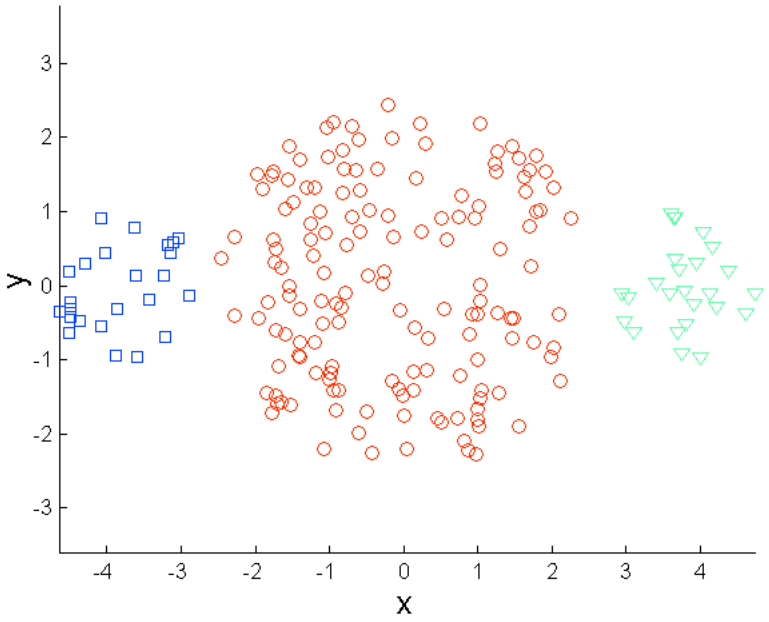
- Ένας από τους πιο γνωστούς αλγόριθμους συσταδοποίησης αυτής της κατηγορίας είναι **ο αλγόριθμος των Κ-μέσων** (K-means).
- Ο **αριθμός Κ** των συστάδων **καθορίζεται πριν** την εκτέλεση του αλγορίθμου.
- Ο αλγόριθμος ξεκινά διαλέγοντας **Κ τυχαία σημεία** από τα δεδομένα ως τα **κέντρα των συστάδων**.
- Έπειτα αναθέτει κάθε σημείο στην συστάδα της οποίας το κέντρο είναι πιο κοντά (μικρότερη απόσταση) σε αυτό το σημείο.
- Στη συνέχεια, υπολογίζει για κάθε συστάδα το **μέσο όρο** όλων των σημείων της (μέσο διάνυσμα) και **ορίζει αυτό ως νέο κέντρο της**.
  - Τα 2 τελευταία βήματα **επαναλαμβάνονται** για ένα προκαθορισμένο αριθμό βημάτων ή μέχρι να μην υπάρχει αλλαγή στο διαχωρισμό των σημείων σε συστάδες.

k-means

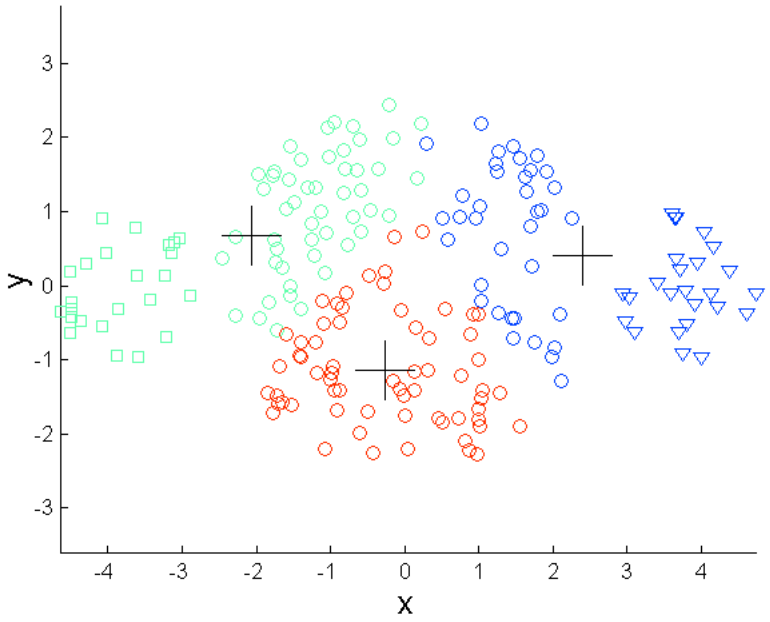
- Επιλογή αρχικών σημείων: πολύ σημαντική!

- Προβλήματα
  - Όταν οι συστάδες έχουν:
    - **Διαφορετικά Μεγέθη**
    - **Διαφορετικές Πυκνότητες**
    - **Μη-σφαιρικά** σχήματα (Non-globular shapes)
  - Όταν τα δεδομένα έχουν **outliers**

K-means: Περιορισμοί – διαφορετικά μεγέθη



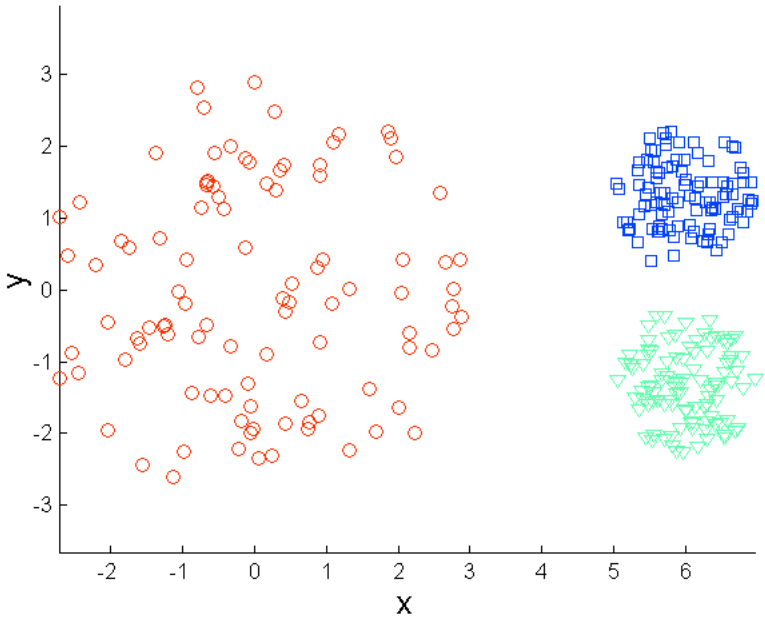
Αρχικά σημεία



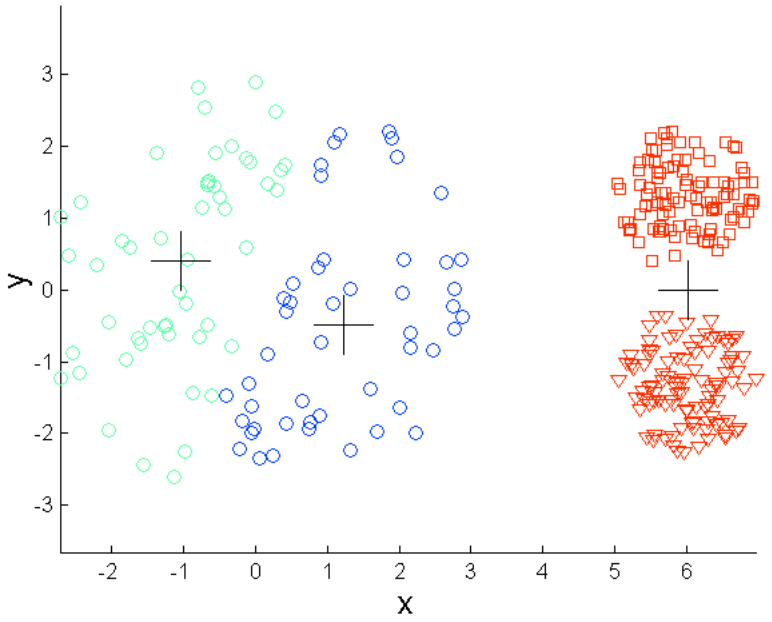
K-means (3 συστάδες)

Δεν μπορεί να βρει το μεγάλο κόκκινο, γιατί είναι πολύ μεγαλύτερο από τα άλλα!

K-means: Περιορισμοί – διαφορετικές πυκνότητες



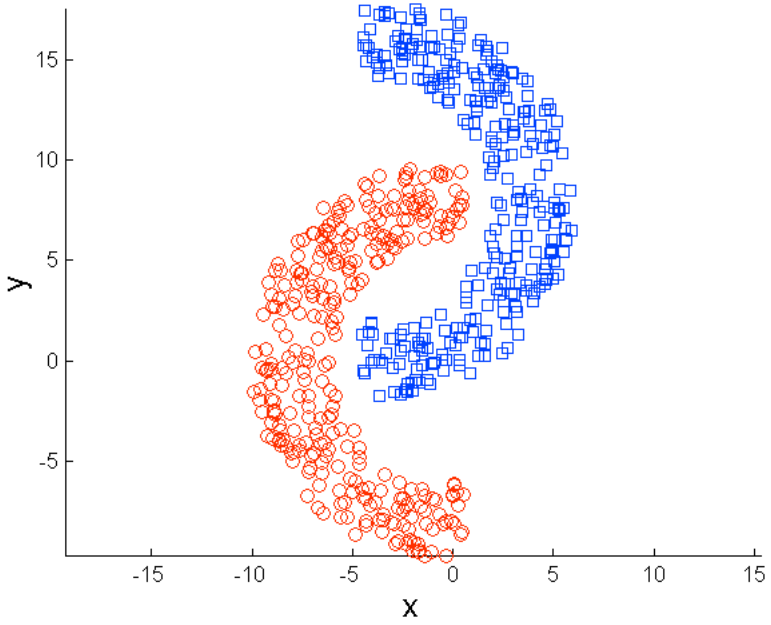
Αρχικά σημεία



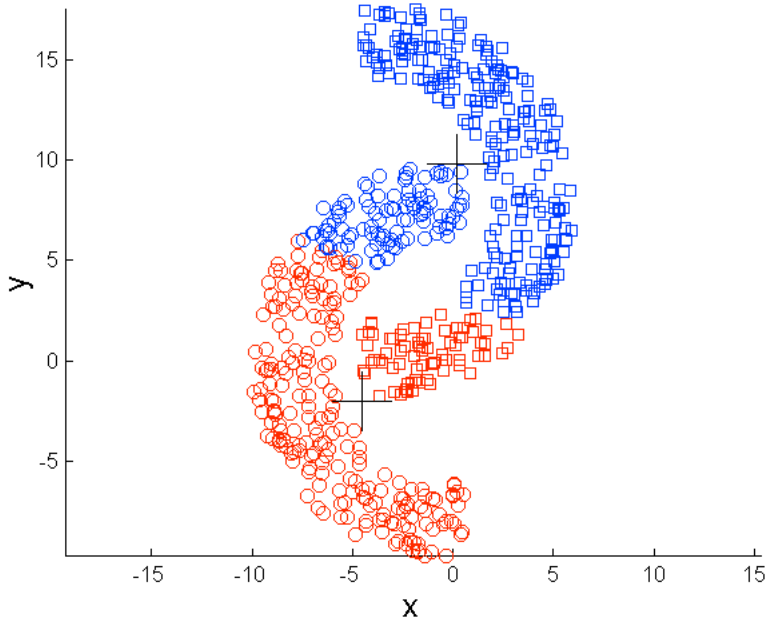
K-means (3 συστάδες)

Δεν μπορεί να διαχωρίσει τα δύο μικρά, γιατί είναι πολύ πυκνά σε σχέση με το ένα μεγάλο!

K-means: Περιορισμοί – μη σφαιρικά σχήματα



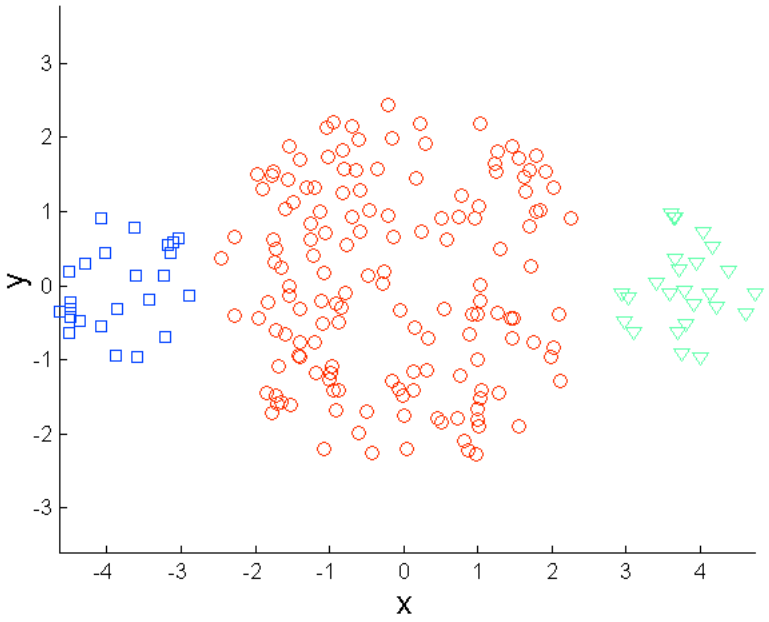
Αρχικά σημεία



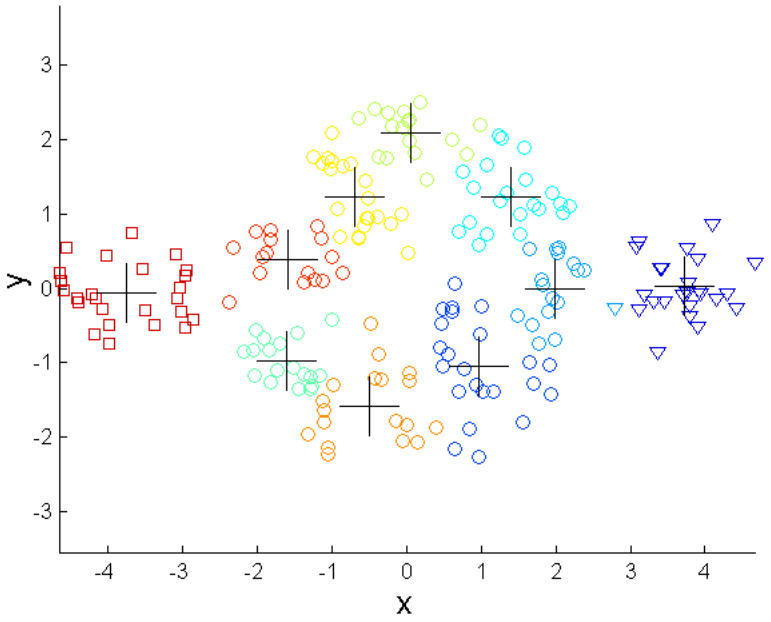
K-means (2 συστάδες)

Δεν μπορεί να βρει τις δύο συστάδες, γιατί έχουν μη σφαιρικά σχήματα!

K-means: Περιορισμοί



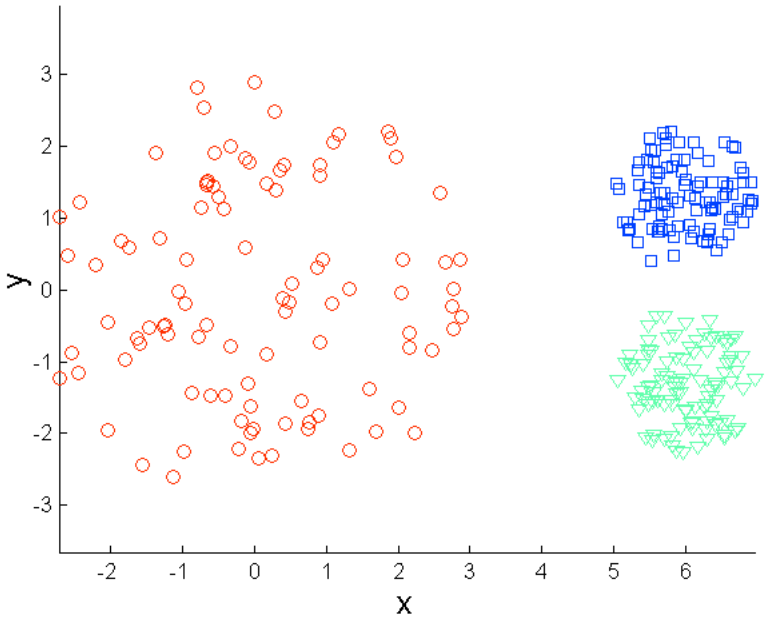
Original Points



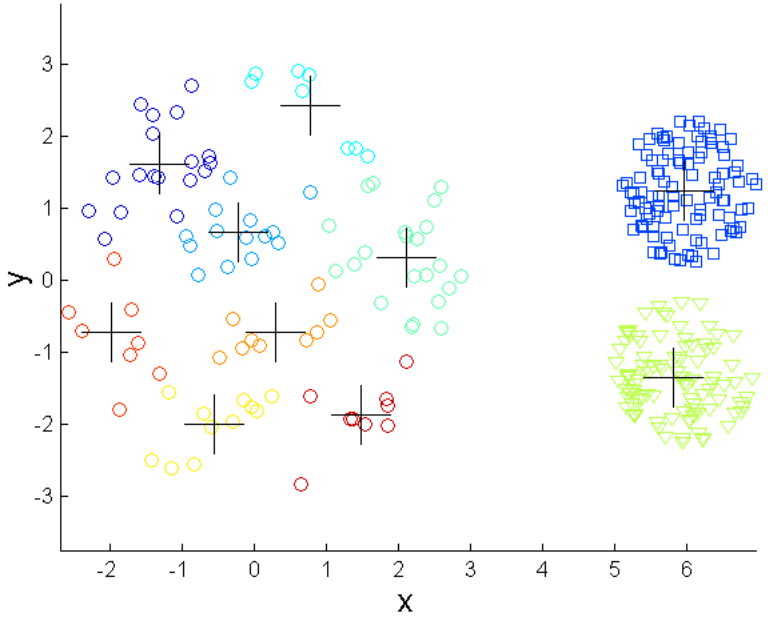
K-means Clusters

Μια λύση είναι να χρησιμοποιηθούν πολλές συστάδες.
Βρίσκει τμήματα των συστάδων, αλλά πρέπει να τα συγκεντρώσουμε.

K-means: Περιορισμοί

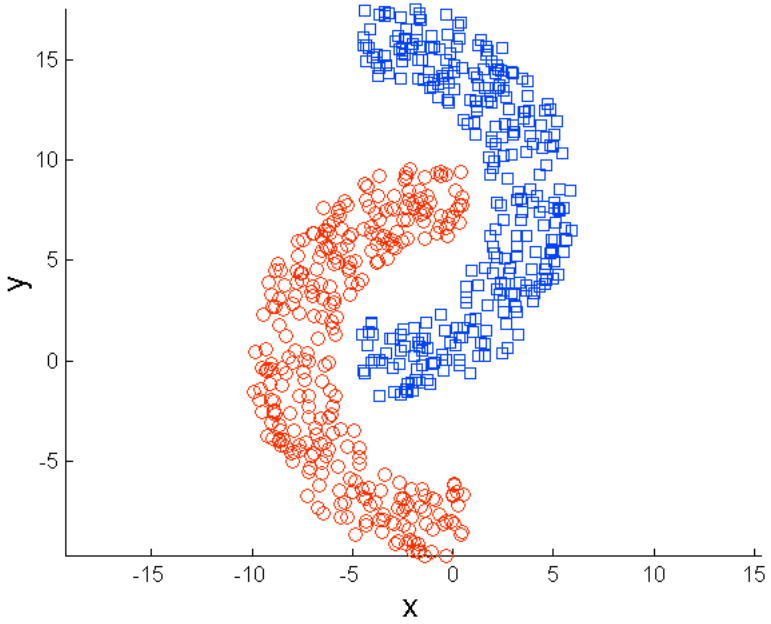


Αρχικά σημεία

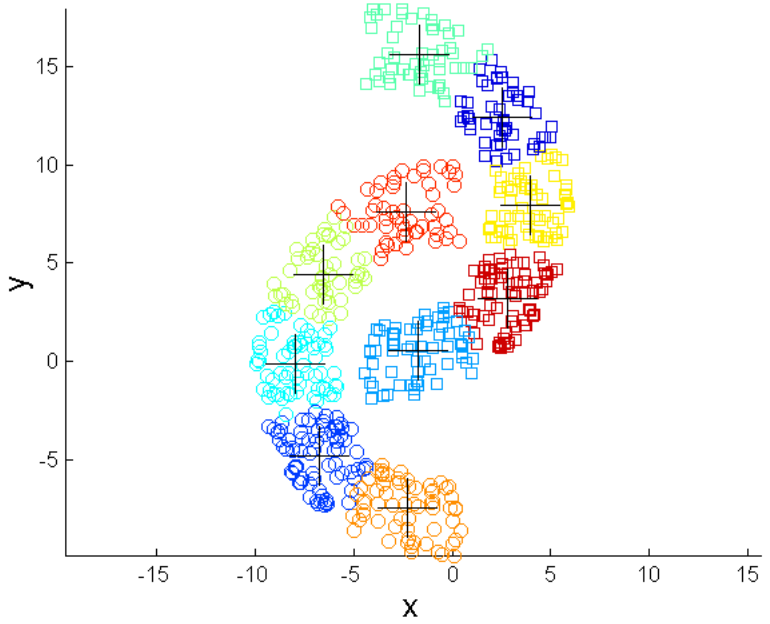


K-means Συστάδες

K-means: Περιορισμοί – διαφορετικά μεγέθη



Αρχικά Σημεία



K-means Συστάδες

Fuzzy c-means

- Αποτελεί **επέκταση του k-means**
- Εν γένει ο ιεραρχικός **k-means** δημιουργεί διαχωρισμούς (partitions)
  - κάθε σημείο δεδομένων μπορεί να **ανήκει σε ένα μόνο cluster**
- Ο **fuzzy c-means** επιτρέπει στα σημεία των δεδομένων να ανήκουν σε περισσότερα από ένα clusters!
  - κάθε σημείο δεδομένων έχει ένα **βαθμό συμμετοχής** (**degree of membership**), ή αλλιώς μία **πιθανότητα** να συμμετέχει σε ένα cluster

Fuzzy c-means

- Έστω ότι το x_i είναι ένα διάνυσμα τιμών για το σημείο δεδομένων g_i.

1. Αρχικοποιούμε την συμμετοχή $U^{(0)} = [\ u_{ij}\ ]$ για το σημείο δεδομένων g_i του cluster cl_j με τυχαίο τρόπο
2. Στο k-ιοστό βήμα, υπολογίζουμε το fuzzy centroid $C^{(k)} = [\ c_j\ ]$ για j = 1, .., nc, όπου το nc είναι ο αριθμός των clusters, χρησιμοποιώντας την:

$$c_j = \frac{\sum_{i=1}^{n} (u_{ij})^m x_i}{\sum_{i=1}^{n} (u_{ij})^m}$$

όπου m είναι η fuzzy παράμετρος και n είναι ο αριθμός των σημείων δεδομένων.

Fuzzy c-means

3. Ενημερώνουμε τη fuzzy membership $U^{(k)} = [\, u_{ij} \,]$, χρησιμοποιώντας την σχέση:

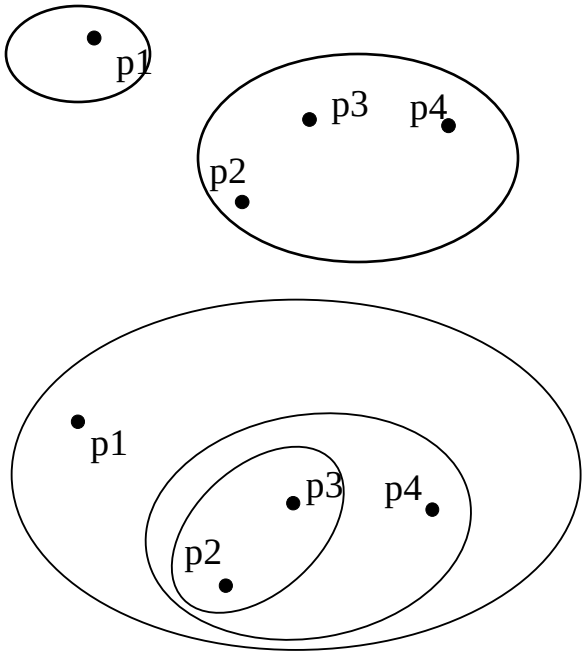
$$u_{ij} = \frac{\left(\frac{1}{\left\| x_i - c_j \right\|} \right)^{\frac{1}{(m-1)}}}{\sum_{j=1}^{n_c} \left(\frac{1}{\left\| x_i - c_j \right\|} \right)^{\frac{1}{(m-1)}}}$$

4. Αν $||U^{(k)} - U^{(k-1)}|| < \varepsilon$, τότε ΣΤΑΜΑΤΑΜΕ, διαφορετικά επιστρέφουμε στο βήμα 2.
5. Αποφασίζουμε το membership cutoff
   - Για κάθε σημείο δεδομένων g_i, τοποθετούμε το g_i στο cluster cl_j αν u_{ij} του $U^{(k)} > \alpha$

Fuzzy c-means

- **Πλεονεκτήματα**
  - Επιτρέπει σε **ένα σημείο να ανήκει σε πολλαπλά clusters**
  - Αποτελεί μια **πιο φυσική αναπαράσταση** της συμπεριφοράς των αληθινών δεδομένων
    - συνήθως εμπλέκονται σε περισσότερες από μία περιπτώσεις
- **Μειονεκτήματα**
  - Πρέπει να καθοριστεί το c, ο αριθμός των clusters
  - Πρέπει να αποφασιστεί η τιμή του membership cutoff
  - Τα clusters είναι ευαίσθητα στην αρχική κατανομή των centroids
  - Ο Fuzzy c-means δεν είναι ντετερμινιστικός αλγόριθμος

Διαχωριστική και Ιεραρχική Συσταδοποίηση



Διαχωριστική Συσταδοποίηση

Ιεραρχική Συσταδοποίηση

Β. Αλγόριθμοι Ιεραρχικής Συσταδοποίησης

- Οι αλγόριθμοι ιεραρχικής συσταδοποίησης **συνδυάζουν συστάδες σε μεγαλύτερες** συστάδες **ή διαιρούν μεγάλες συστάδες** σε μικρότερες.
- Το αποτέλεσμα των αλγορίθμων αυτών είναι μια ιεραρχία από διαφορετικές ομαδοποιήσεις των δεδομένων στο ένα άκρο της οποίας βρίσκεται μια μόνο συστάδα με όλα τα δεδομένα, και στο άλλο τόσες συστάδες όσες και ο αριθμός των δεδομένων.
- Με βάση την **κατεύθυνση ανάπτυξης** της ιεραρχίας που ακολουθούν, οι ιεραρχικοί αλγόριθμοι ομαδοποίησης χωρίζονται στους **αλγορίθμους συγχώνευσης** (**agglomerative**) και στους **αλγορίθμους διαίρεσης** (**divisive**).
  - Οι αλγόριθμοι **συγχώνευσης** είναι οι πιο σημαντικοί και διαδεδομένοι από τους δύο. Βασίζονται σε μετρικές απόστασης ανάμεσα σε συστάδες.
  - Δεδομένης μιας αρχικής συσταδοποίησης (π.χ. κάθε σημείο αποτελεί μια συστάδα), οι αλγόριθμοι αυτοί βρίσκουν τις δύο πιο κοντινές συστάδες και τις συγχωνεύουν με μία.
  - Η διαδικασία συνεχίζεται μέχρις ότου προκύψει **μία μόνο συστάδα**!

Β. Αλγόριθμοι Ιεραρχικής Συσταδοποίησης

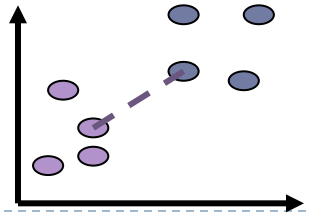
- Αλγόριθμος:
  - Έστω σημεία ενός συνόλου S
  - Τοποθετήστε κάθε σημείο του S σε ένα δικό του cluster (singleton), δημιουργώντας μία λίστα από clusters L:

$$L = S_1, S_2, S_3, ..., S_{n-1}, S_n$$

  - Υπολογίστε μια **συνάρτηση κόστους συγχώνευσης** (merging cost function) μεταξύ κάθε ζεύγους των στοιχείων στη L για να βρείτε τις δύο πιο κοντινές συστάδες $\{S_i, S_j\}$, το οποίο θα είναι το φθηνότερο ζευγάρι προς συγχώνευση.
  - Αφαιρέστε τα S_i και S_j από τη L.
  - Συγχωνεύστε τα S_i και S_j για να δημιουργήσετε ένα νέο εσωτερικό κόμβο S_{ij} στη T, ο οποίος θα είναι ο γονιός των S_i και S_j στο προκύπτον δέντρο.
  - Πηγαίνετε στο βήμα 2 μέχρι να υπάρχει μόνο 1 σύνολο που απομένει.

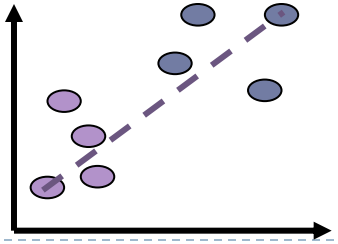
Β. Αλγόριθμοι Ιεραρχικής Συσταδοποίησης

- Το Βήμα 2 μπορεί να γίνει με διάφορους τρόπους, κάτι που διακρίνει την συσταδοποίηση σε:
  - single-linkage
  - complete-linkage
  - average-linkage
- Στη **single-linkage συσταδοποίηση** θεωρούμε ότι η απόσταση μεταξύ ενός cluster και ενός άλλου cluster είναι ίση με την **μικρότερη** απόσταση από κάθε μέλος ενός cluster σε οποιοδήποτε μέλος του άλλου cluster.



Β. Αλγόριθμοι Ιεραρχικής Συσταδοποίησης

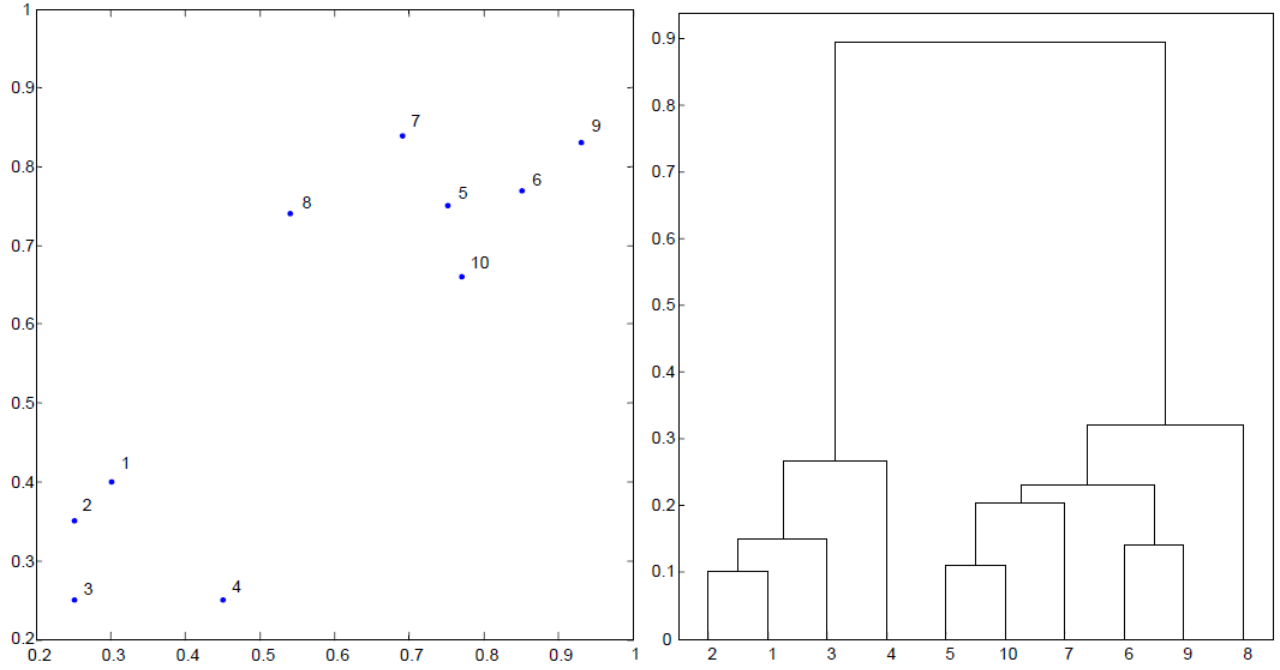
- Το Βήμα 2 μπορεί να γίνει με διάφορους τρόπους, κάτι που διακρίνει την συσταδοποίηση σε:
  - single-linkage
  - complete-linkage
  - average-linkage
- Στην **complete-linkage συσταδοποίηση** θεωρούμε ότι η απόσταση μεταξύ ενός cluster και ενός άλλου cluster είναι ίση με τη **μεγαλύτερη** απόσταση από οποιοδήποτε μέλος ενός cluster σε οποιοδήποτε μέλος του άλλου cluster.



Β. Αλγόριθμοι Ιεραρχικής Συσταδοποίησης

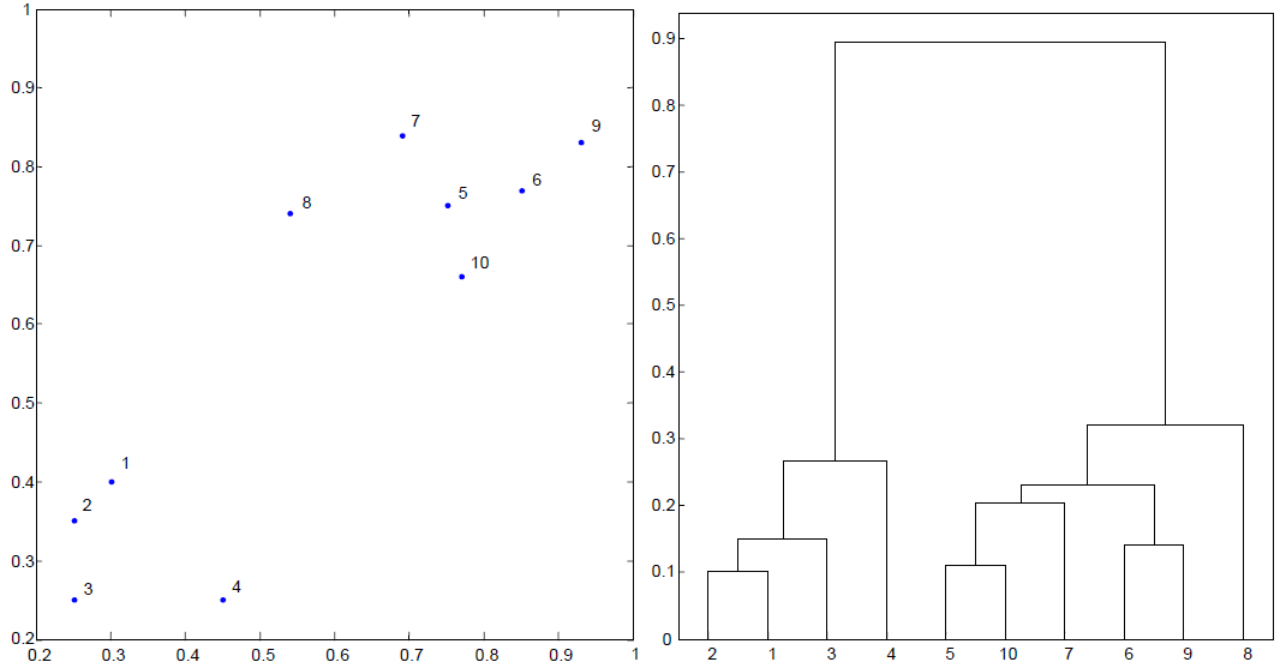
- Το Βήμα 2 μπορεί να γίνει με διάφορους τρόπους, κάτι που διακρίνει την συσταδοποίηση σε:
  - single-linkage
  - complete-linkage
  - average-linkage
- Στην **average-linkage** συσταδοποίηση θεωρούμε ότι η απόσταση μεταξύ ενός cluster και ενός άλλου cluster είναι **ίση με τη μέση απόσταση** από οποιοδήποτε μέλος ενός cluster σε οποιοδήποτε μέλος του άλλου cluster.

- Οι ιεραρχίες που προκύπτουν από τους αλγορίθμους ιεραρχικής ομαδοποίησης μπορεί να απεικονιστούν με έναν πρακτικό και εύκολο τρόπο μέσω ενός **γραφήματος δενδρικής μορφής**, το οποίο ονομάζεται **δενδρόγραμμα**.



Δενδρόγραμμα Ιεραρχικής Ομαδοποίησης

- Το τελικό cluster είναι η ρίζα και κάθε δεδομένο είναι ένα φύλλο.
- Το ύψος των επιμέρους «ράβδων» καταδεικνύει πόσο «κοντά» είναι τα αντικείμενα

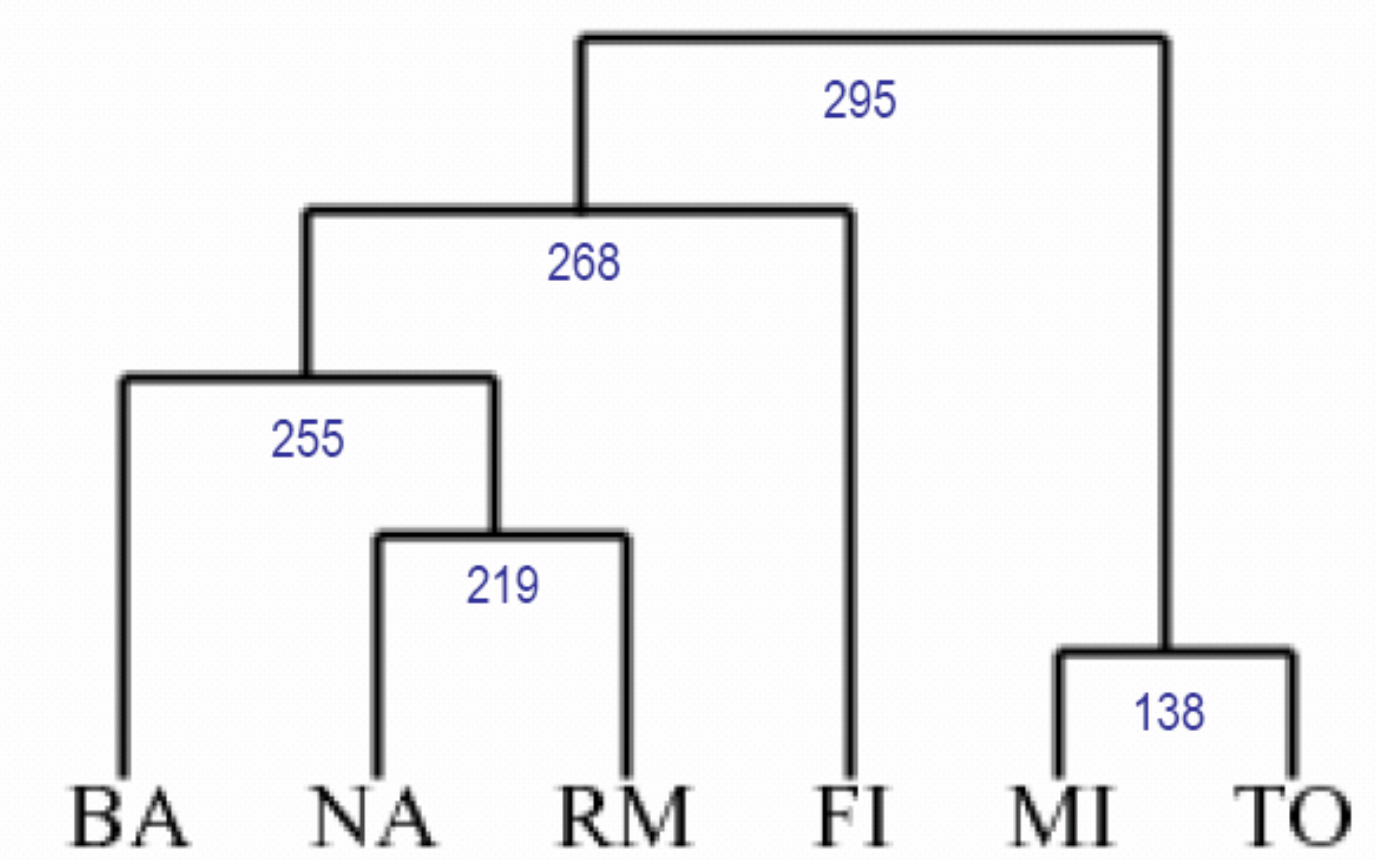


Δενδρόγραμμα Ιεραρχικής Ομαδοποίησης

Παράδειγμα Ιεραρχικής Συσταδοποίησης

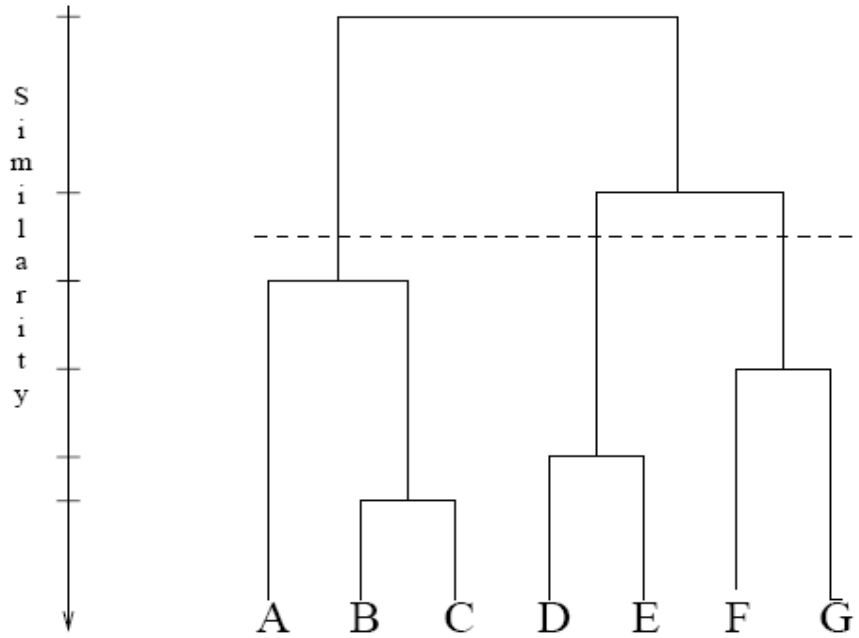
	BA	FI	MI	NA	RM	TO
BA	0	662	877	255	412	996
FI	662	0	295	468	268	400
MI	877	295	0	754	564	138
NA	255	468	754	0	219	869
RM	412	268	564	219	0	669
TO	996	400	138	869	669	0

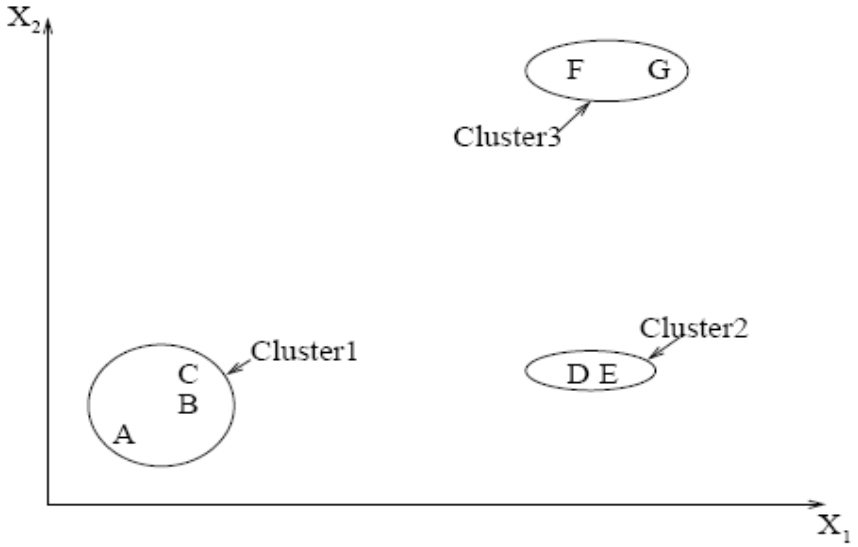
Ιεραρχική Συσταδοποίηση: χρήση single linkage



Ιεραρχική Συσταδοποίηση: δημιουργία clusters

- Forming clusters from dendograms





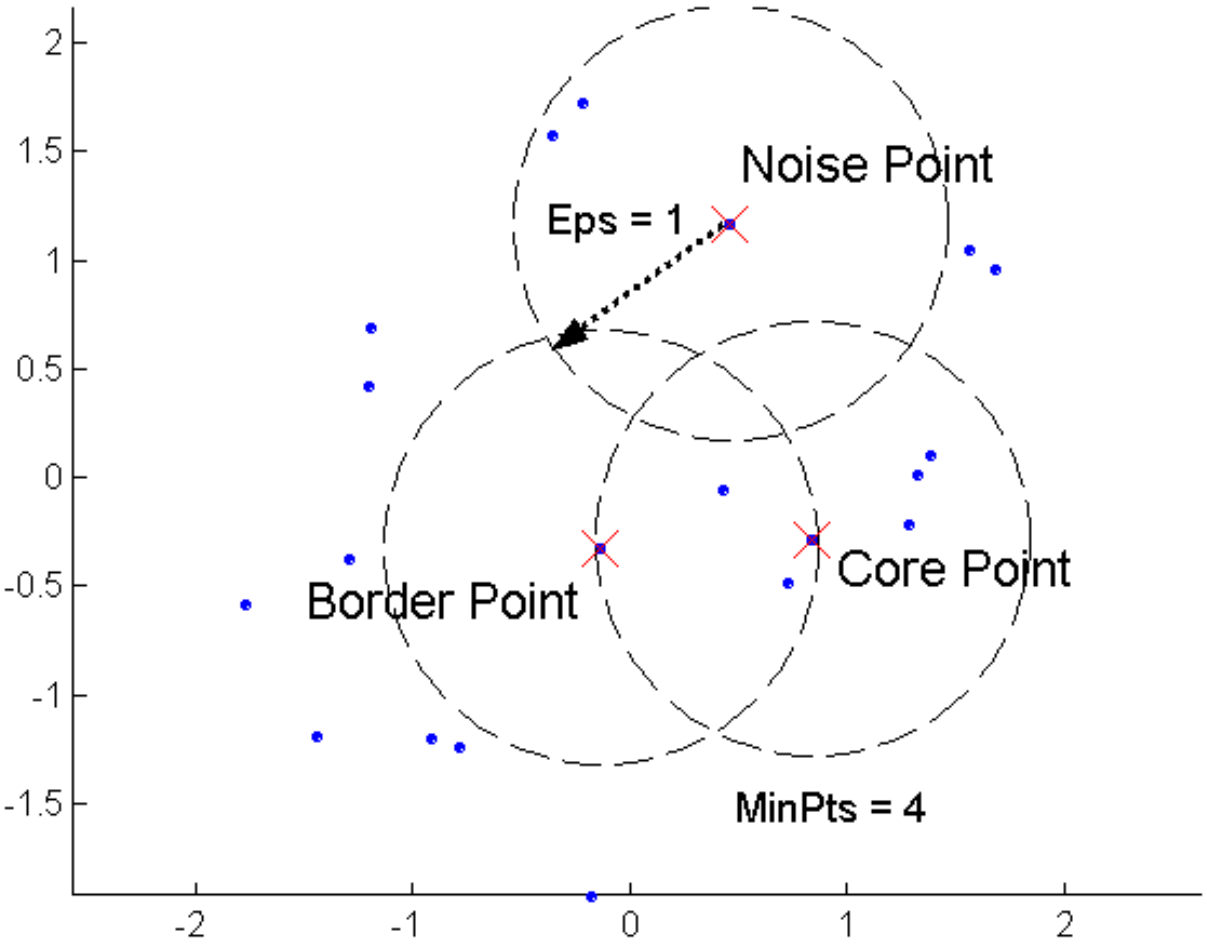
Ιεραρχική Συσταδοποίηση

- Πλεονεκτήματα
  - Τα δενδρογράμματα είναι ιδανικά για οπτικοποίηση (visualization)
  - Παρέχει ιεραρχικές σχέσεις μεταξύ των clusters
  - Φαίνεται να είναι σε θέση να συλλάβει ομόκεντρα clusters
- Μειονεκτήματα
  - Δεν είναι εύκολο να καθοριστούν τα επίπεδα του clustering
  - Πειράματα έδειξαν ότι άλλες τεχνικές συσταδοποίησης ξεπερνούν την απόδοση της ιεραρχικής συσταδοποίησης.

DBSCAN

- DBSCAN: Density-based spatial clustering of applications with noise
- Ο DBSCAN είναι ένας αλγόριθμος βασισμένος στην πυκνότητα
- **Πυκνότητα** = αριθμός σημείων μέσα σε μια προκαθορισμένη ακτίνα (Eps)
- Τα σημεία διαχωρίζονται σε:
  - **Βασικά** (core): ένα σημείο για το οποίο υπάρχουν **περισσότερα από ένα προκαθορισμένο αριθμό (MinPts) σημεία** σε ακτίνα Eps
    - Αυτά είναι τα σημεία που είναι στο εσωτερικό μιας συστάδας
  - **Οριακά** (border): ένα σημείο για το οποίο υπάρχουν **λιγότερα από ένα προκαθορισμένο αριθμό (MinPts) σημεία** σε ακτίνα Eps, αλλά είναι στη γειτονιά ενός βασικού σημείου
  - **Θορύβου** (noise): ένα σημείο που **δεν είναι ούτε βασικό, ούτε οριακό**

DBSCAN: Γενικά

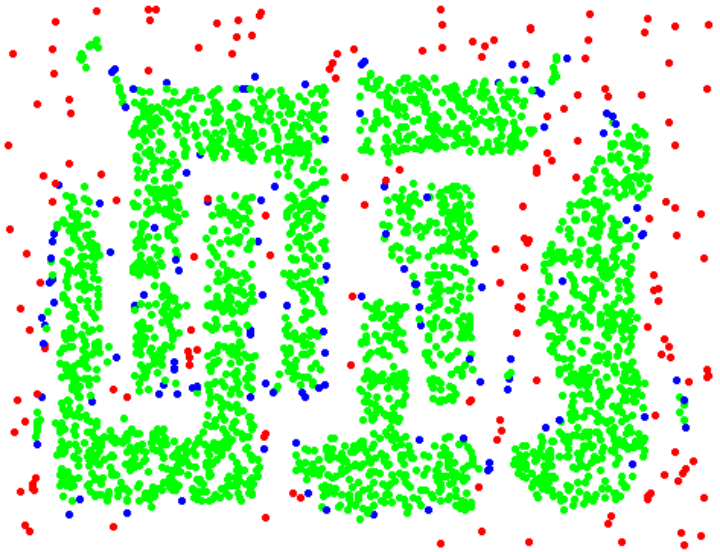


DBSCAN: Αλγόριθμος

1. Χαρακτήρισε κάθε σημείο ως **βασικό**, **οριακό** ή **θόρυβο**
2. Διέγραψε τα σημεία θορύβου
3. Τοποθέτησε μια ακμή μεταξύ όλων των **βασικών** σημείων που είναι **σε απόσταση έως Eps μεταξύ τους**
4. Κάνε κάθε ομάδα συνδεδεμένων βασικών σημείων μια διαφορετική **συστάδα**
5. Ανάθεσε **κάθε οριακό σημείο** σε μία από τις συστάδες των συσχετιζόμενων του βασικών σημείων

DBSCAN: Αλγόριθμος

Αρχικά σημεία

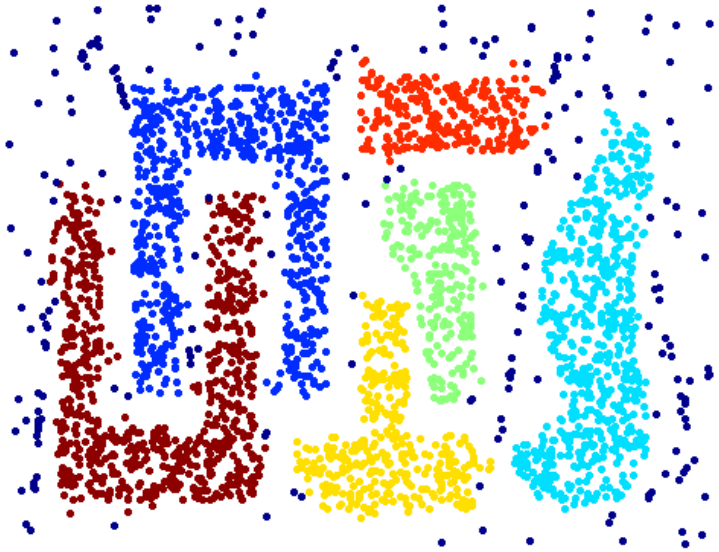


Τύποι σημείων: core, border και noise

Eps = 10, MinPts = 4

DBSCAN: Πλεονέκτημα

Αρχικά Σημεία



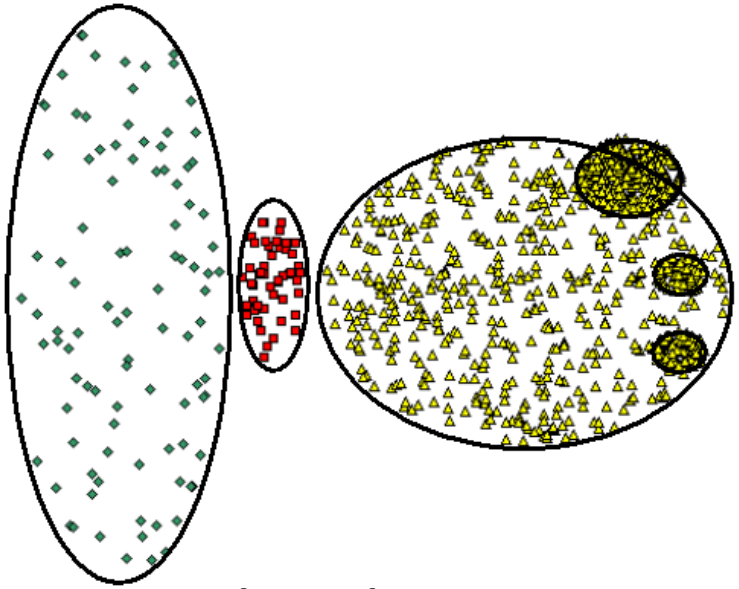
Συστάδες

- Δεν επηρεάζεται από το θόρυβο
- Μπορεί να χειριστεί συστάδες με διαφορετικά σχήματα και μεγέθη

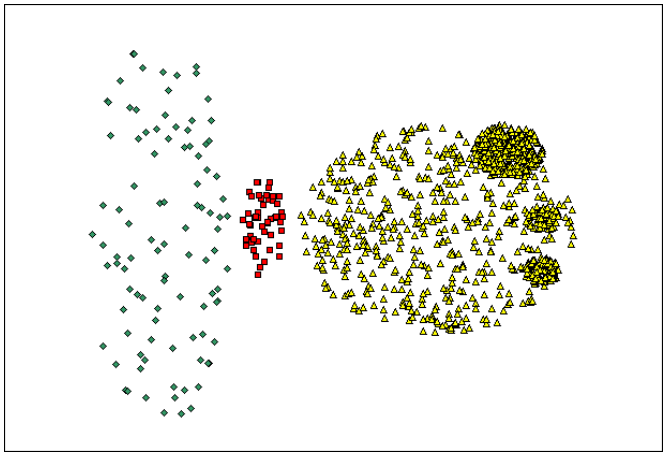
DBSCAN: Πολυπλοκότητα

- O(m x χρόνος εντοπισμού σημείων σε eps-γειτονιά)
  - $O(m^2)$
  - Για μικρό αριθμό διαστάσεων, υπάρχουν δομές που υποστηρίζουν την πράξη σε O(m logm)
- O(m) χώρος (κρατάμε μόνο ένα label)

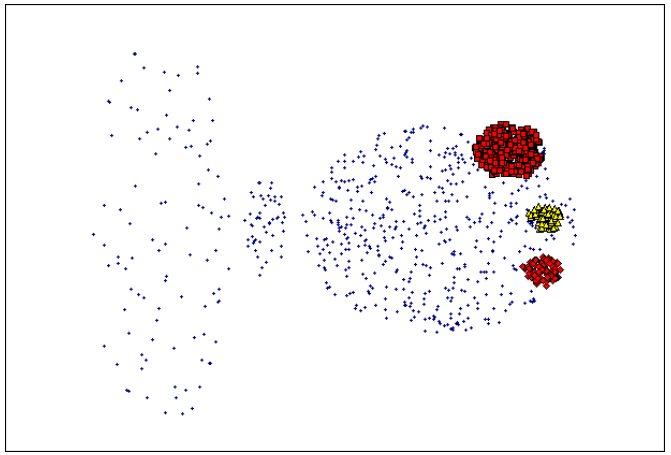
DBSCAN: Περιορισμοί



Αρχικά Σημεία



(MinPts=4, Eps=9.75).



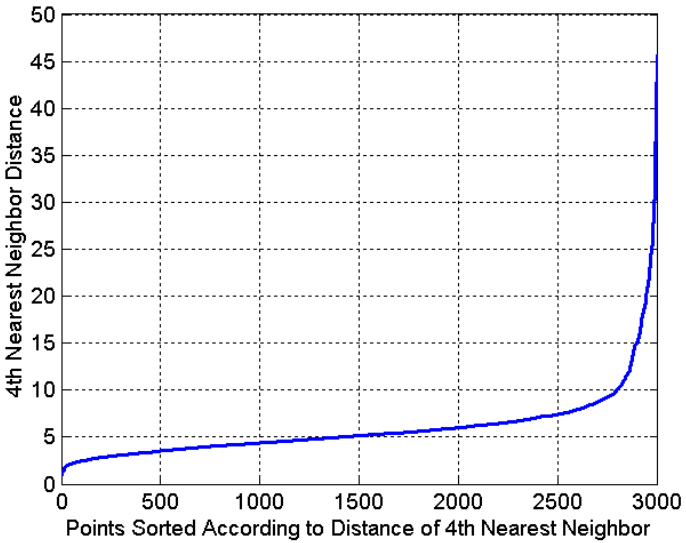
(MinPts=4, Eps=9.92)

- **Διαφορετικές πυκνότητες**
- **Πολυ-διάστατα δεδομένα – δύσκολος ορισμός πυκνότητας και δαπανηρός υπολογισμός γειτόνων**

DBSCAN: Καθορισμός των MinPts και Eps

- Η ιδέα είναι να κοιτάξουμε την απόσταση ενός σημείου από τον k-οστό κοντινότερο γείτονα του -> k-dist
- Γενικά, για τα σημεία που ανήκουν στην ίδια συστάδα, η τιμή του k-dist θα είναι μικρή (αν το k δεν είναι μεγαλύτερο από το μέγεθος της συστάδας)
- Θα θέλαμε για τα σημεία μιας συστάδας, να έχουν περίπου την ίδια k-dist
- Τα σημεία θορύβου έχουν μεγαλύτερες k-dist

- Υπολογίζουμε την k-dist για όλα τα σημεία, για κάποιο k
- Ταξινομούμε τις αποστάσεις με φθίνουσα διάταξη
- Περιμένουμε ξαφνική αλλαγή στο k-dist που αντιστοιχεί στο Eps
- Οπότε k = MinPts και Eps = k-dist



DBSCAN: **Πλεονεκτήματα**

- Σε αντίθεση με αλγορίθμους, όπως ο k-means, ο DBSCAN **δεν απαιτεί τον εκ των προτέρων προσδιορισμό του αριθμού των συστάδων**.
- Μπορεί να **καταλήξει σε αυθαίρετα σχήματα συστάδων**. Μπορεί να εντοπίσει ακόμα και μια συστάδα, η οποία βρίσκεται γύρω από κάποια άλλη.
  - Αυτό συμβαίνει λόγω της παραμέτρου MinPts, η οποία ελαττώνει την εμφάνιση του φαινομένου της αλυσίδας συστάδων. Το φαινόμενο της αλυσίδας συστάδων συμβαίνει, όταν διαφορετικές συστάδες συνδέονται με μια λεπτή γραμμή σημείων-αντικειμένων.
- Έχει **καλή ευαισθησία στον θόρυβο** και δεν επηρεάζεται από ακραίες τιμές.
- Χρειάζεται **μόνο δύο παραμέτρους** και έχει μικρή ευαισθησία ως προς τη σειρά εμφάνισης των δεδομένων στη βάση.
- Εφόσον έχουν μελετηθεί τα δεδομένα και έχουν γίνει κατανοητά, ο προσδιορισμός των παραμέτρων MinPts και ε δεν είναι δύσκολος.

DBSCAN: **Μειονεκτήματα**

- Δεν είναι **απόλυτα ντετερμινιστικός**, υπό την έννοια ότι τα περιθωριακά σημεία μιας συστάδας μπορούν να ανήκουν είτε σε αυτή είτε σε κάποια γειτονική, ανάλογα με τη σειρά επεξεργασίας.
  - Ευτυχώς, αυτό δεν συμβαίνει συχνά και έχει μικρό αντίκτυπο στα αποτελέσματα.
- Η ποιότητα των αποτελεσμάτων **εξαρτάται από τη μετρική απόστασης** που θα χρησιμοποιηθεί.
  - Η πιο κοινή μετρική απόστασης είναι η **Ευκλείδεια απόσταση**. Όμως, ειδικά για πολυδιάστατα δεδομένα, η συγκεκριμένη μετρική είναι **σχεδόν άχρηστη(!)**, λόγω της λεγόμενης **«κατάρας της διαστατικότητας»**, κάνοντας έτσι δύσκολη την επιλογή της παραμέτρου ε. Ωστόσο, αυτό μπορεί να συμβεί σε οποιονδήποτε αλγόριθμο χρησιμοποιεί την Ευκλείδεια απόσταση.
- Δεν μπορεί να συσταδοποιήσει καλά σύνολα από δεδομένα με μεγάλες διαφορές πυκνότητας, καθώς δεν μπορεί να εντοπιστεί κάποιος συνδυασμός MinPts-ε, που να είναι κατάλληλος για όλες τις συστάδες.
- Αν τα δεδομένα δεν έχουν γίνει κατανοητά, η επιλογή ενός κατωφλίου ε που να έχει νόημα μπορεί να είναι δύσκολη.

Περιεχόμενα Β' μέρους

- Γενετικοί αλγόριθμοι
  - Ιστορικά στοιχεία
  - Μηχανισμός λειτουργίας
  - Συστατικά Γενετικού Αλγορίθμου
- Νευρωνικά Δίκτυα
  - Εισαγωγικά στοιχεία
  - Συναρτήσεις Ενεργοποίησης
  - Τεχνητά Νευρωνικά Δίκτυα
  - ΤΝΔ Πρόσθιας Τροφοδότησης, Perceptron, Μνήμες Συσχέτισης
  - Εφαρμογές Νευρωνικών Δικτύων

Γενετικοί Αλγόριθμοι

- Σε αρκετές περιπτώσεις το μέγεθος ενός προβλήματος καθιστά **απαγορευτική** τη **χρήση κλασικών μεθόδων αναζήτησης** για την επίλυσή του. Από την άλλη τυχαίνει να είναι εύκολο να δημιουργηθούν λύσεις με απευθείας μηχανισμούς.
  - π.χ. στο πρόβλημα του πλανόδιου πωλητή (TSP) ο οποιοσδήποτε μπορεί να ορίσει μια πορεία που περνά από όλες τις πόλεις μία φορά – πλην όμως μάλλον δεν θα είναι η καλύτερη δυνατή!
- Στις περιπτώσεις αυτές βρίσκουν εφαρμογή **πιθανοκρατικοί αλγόριθμοι** οι οποίοι **αν και δεν εγγυώνται ότι θα βρουν τη βέλτιστη λύση**, είναι ικανοί να επιστρέψουν **μια αρκετά καλή λύση** σε εύλογο χρονικό διάστημα.
- Μια κατηγορία τέτοιων αλγορίθμων επίλυσης προβλημάτων είναι οι **γενετικοί αλγόριθμοι ή ΓΑ (genetic algorithms),** στους οποίους ο βασικός μηχανισμός λειτουργίας είναι εμπνευσμένος από τη Δαρβινική Θεωρία της Εξέλιξης.

Γενετικοί Αλγόριθμοι

- Σε αντίθεση με τους κλασικούς αλγόριθμους αναζήτησης που ψάχνουν στο χώρο των καταστάσεων, οι ΓΑ **εκτελούν αναζήτηση στο χώρο των υποψήφιων λύσεων με στόχο την εύρεση αποδεκτών λύσεων, σύμφωνα με κάποιο κριτήριο**.
  - να γιατί πρέπει να είναι εύκολο να δημιουργηθούν λύσεις με απευθείας μηχανισμούς!

Ιστορικά Στοιχεία ΓΑ

- Το 1958 ο Friedberg, επιχείρησε να συνδυάσει μικρά προγράμματα FORTRAN, ωστόσο τα προγράμματα που προέκυψαν συνήθως δεν ήταν εκτελέσιμα.

- Το 1975 ο Holland, χρησιμοποίησε σειρές bits για να αναπαραστήσει λειτουργίες με τρόπο τέτοιο, ώστε κάθε συνδυασμός bits να είναι μια έγκυρη λειτουργία.

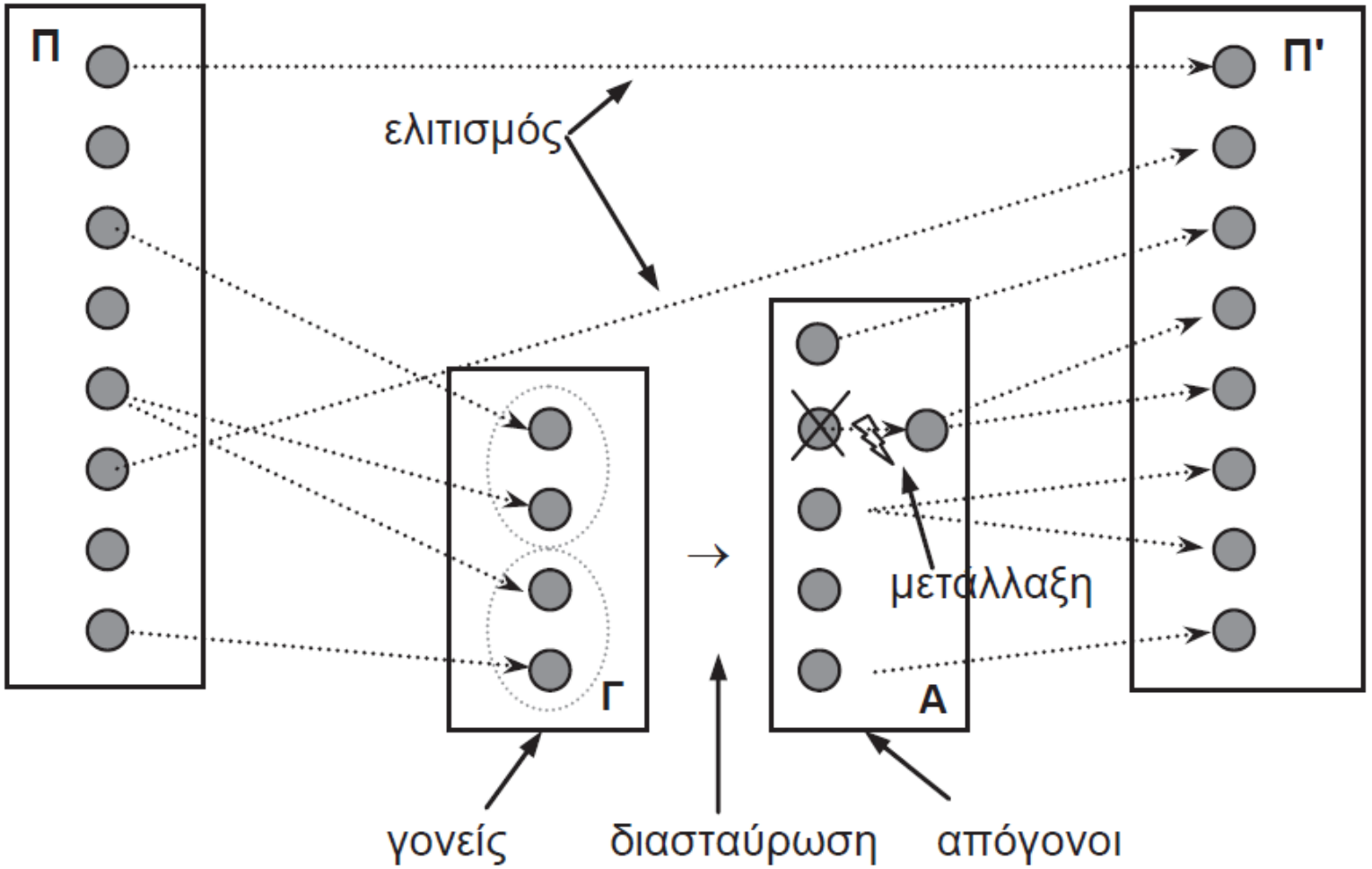
Γενετικοί Αλγόριθμοι

- Στην επιστήμη των υπολογιστών ένας Γενετικός Αλγόριθμος (ΓΑ) είναι ένα metaheuristic που εμπνέεται από τη **διαδικασία της φυσικής επιλογής** και ανήκει στη μεγαλύτερη τάξη εξελικτικών αλγορίθμων (ΕΑ).
- Οι Γενετικοί Αλγόριθμοι χρησιμοποιούνται συνήθως για τη δημιουργία λύσεων υψηλής ποιότητας σε **προβλήματα βελτιστοποίησης** και **αναζήτησης**, βασιζόμενοι σε διαδικασίες εμπνευσμένες από τη βιολογία (bio-inspired), όπως είναι η **μετάλλαξη** (mutation), η **διασταύρωση** (crossover) και η **επιλογή** (selection).

Θεωρία της Εξέλιξης (evolution)

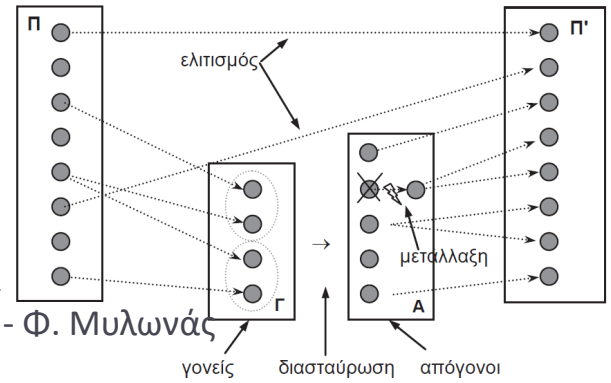
- **Κανόνας της φυσικής επιλογής**:
  - Οι οργανισμοί που **δε μπορούν να επιβιώσουν** στο περιβάλλον τους **πεθαίνουν**, ενώ οι υπόλοιποι πολλαπλασιάζονται μέσω της αναπαραγωγής.
  - Οι **απόγονοι** παρουσιάζουν **μικρές διαφοροποιήσεις** από τους **προγόνους** τους, ενώ **συνήθως υπερισχύουν αυτοί που συγκεντρώνουν τα καλύτερα χαρακτηριστικά**.
  - Σποραδικά συμβαίνουν **τυχαίες μεταλλάξεις**, από τις οποίες οι περισσότερες οδηγούν τα μεταλλαγμένα άτομα στο θάνατο, αν και είναι πιθανό, πολύ σπάνια όμως, να οδηγήσουν στη δημιουργία νέων "καλύτερων" οργανισμών.
  - Αν το περιβάλλον μεταβάλλεται με αργούς ρυθμούς, τα διάφορα είδη μπορούν να εξελίσσονται σταδιακά ώστε να προσαρμόζονται σε αυτό.

Γενικός Γενετικός Αλγόριθμος



Μηχανισμός Λειτουργίας ΓΑ

1. Δημιούργησε έναν τυχαίο **αρχικό πληθυσμό Π**, με Ν υποψήφιες λύσεις (δηλαδή μη έγκυρες ή μη βέλτιστες, κλπ). Αυτές αποτελούν τα **άτομα του πληθυσμού**.
2. Βαθμολόγησε κάθε υποψήφια λύση χρησιμοποιώντας μια **συνάρτηση καταλληλότητας** (**fitness function**) (δηλ. πόσο κοντά σε μια αποδεκτή λύση είναι).
3. Σχημάτισε **ζευγάρια γονέων** λαμβάνοντας με στοχαστικό τρόπο άτομα του πληθυσμού Π. Τα άτομα μπορεί να συμμετέχουν καθόλου ή περισσότερες από μία φορές. Δίνεται όπως μεγαλύτερη προτεραιότητα στις πλέον κατάλληλες λύσεις (άτομα) του πληθυσμού.
4. **Κάθε ζευγάρι διασταυρώνεται** (**mates**), δίνοντας νέες λύσεις (**απογόνους-offsprings**).
5. **Πιθανοκρατικά**, κάποιοι από τους απογόνους υφίστανται **μετάλλαξη** (**mutation**).
6. Ο **νέος πληθυσμός Π'** (ίδιου μεγέθους με τον Π) δημιουργείται **επιλέγοντας** με κάποιο συστηματικό τρόπο "**καλά**" **άτομα**, κύρια από το σύνολο των απογόνων και δευτερευόντως αλλά όχι υποχρεωτικά, από τον αρχικό πληθυσμό.
7. Η **διαδικασία επαναλαμβάνεται** για το νέο πληθυσμό Π' μέχρι να πληρείται κάποια **συνθήκη τερματισμού** (π.χ. εύρεση λύσης ικανοποιητικής ποιότητας, μη περαιτέρω βελτίωση των λύσεων, χρονικά όρια, κτλ.)



Συστατικά Γενετικού Αλγορίθμου

- Ένας ΓΑ για δεδομένο πρόβλημα, περιλαμβάνει:
  1. **Αναπαράσταση** (κωδικοποίηση) **Υποψήφιων Λύσεων**
  2. Ορισμό της **Συνάρτησης Καταλληλότητας** (ποιότητας)
  3. **Δημιουργία Αρχικού Πληθυσμού Λύσεων** (συνήθως δημιουργείται τυχαία)
  4. **Μηχανισμό Επιλογής** Γονέων
  5. **Διαδικασία Αναπαραγωγής/Ανασυνδυασμού**
  6. Ορισμό του **Πληθυσμού** της **Επόμενης Γενιάς**.
- Επιπρόσθετα, υπάρχουν και οι **συνθήκες τερματισμού** της αναζήτησης, οι οποίες όμως είναι γενικές (ανεξάρτητες προβλήματος).

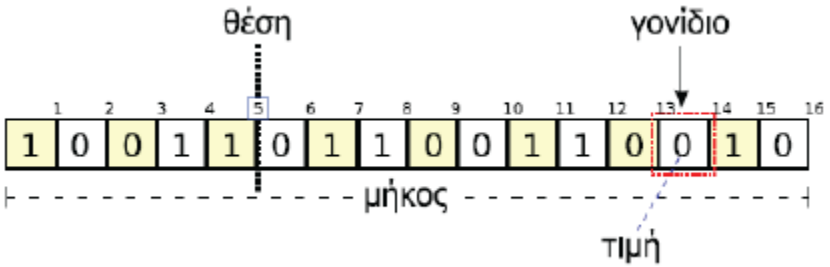
A. Αναπαράσταση Υποψήφιων Λύσεων

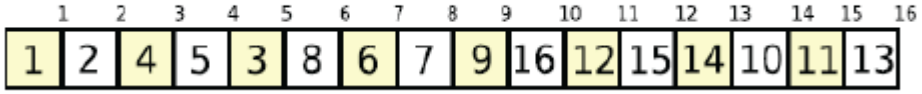
- Στους βιολογικούς οργανισμούς, ένα χρωμόσωμα είναι ένα μεγάλο μόριο DNA και περιέχει έναν αριθμό γονιδίων.
- Το DNA αποτελείται από αλληλουχίες τεσσάρων διαφορετικών νουκλεοτιδίων (βάσεων): **A**denine, **G**uanine, **T**hymine, **C**ytosine.
- Άρα το αλφάβητο του DNA έχει 4 γράμματα: A, G, T και C.

- Στους ΓΑ, **κάθε υποψήφια λύση αναπαρίσταται** με μία **συμβολοσειρά ενός πεπερασμένου αλφάβητου**.
- Η **συμβολοσειρά** αποκαλείται και **χρωμόσωμα** (chromosome) ενώ τα **επιμέρους τμήματά** της που κωδικοποιούν κάποιο χαρακτηριστικό ονομάζονται και **γονίδια** (gene).

Τυπικές Αναπαραστάσεις (μη πλήρης λίστα)

- **Δυαδική Αναπαράσταση**: Χρησιμοποιούνται bits με τιμή 0 ή 1. Η ερμηνεία τους εξαρτάται από το πρόβλημα. Στο παράδειγμα, θα μπορούσε να είναι 2 ακέραιοι των 8 bits έκαστος, ή 4 των 4 bits.
- **Αναπαράσταση Permutation (Συνδυασμού)**
  - Χρησιμοποιούνται ακέραιοι, σε κάποια διάταξη, χωρίς όμως να επαναλαμβάνονται.
  - Θα μπορούσε π.χ. κάθε ακέραιος να είναι μια πόλη και όλο το χρωμόσωμα μια διαδρομή (TSP)!





Β. Συνάρτηση Καταλληλότητας (Fitness Function)

- Δέχεται ως **είσοδο** ένα χρωμόσωμα και επιστρέφει έναν αριθμό, που **υποδηλώνει** το **πόσο κατάλληλο (ή "καλό")** είναι το συγκεκριμένο άτομο-λύση. Π.χ.:
  - Αν το χρωμόσωμα κωδικοποιεί μια διαδρομή (TSP), η τιμή καταλληλότητας θα μπορούσε να είναι το μήκος της, ή ο χρόνος για να τη διανύσει κάποιος, ή το κόστος σε απαιτούμενα καύσιμα, ή και συνδυασμός αυτών!
  - Αν το χρωμόσωμα κωδικοποιεί τη θέση που πρέπει να τοποθετηθούν κεραίες κινητής τηλεφωνίας, η τιμή καταλληλότητας θα μπορούσε να είναι το ποσοστό κάλυψης στην περιοχή.
- Η **αξιολόγηση** των λύσεων χρησιμοποιείται:
  - στην **επιλογή γονέων** για τη διαδικασία αναπαραγωγής
  - στην **επιλογή ατόμων** για τον σχηματισμό του **πληθυσμού** της **επόμενης γενιάς**
  - στην **συνθήκη τερματισμού**

Παράδειγμα 1

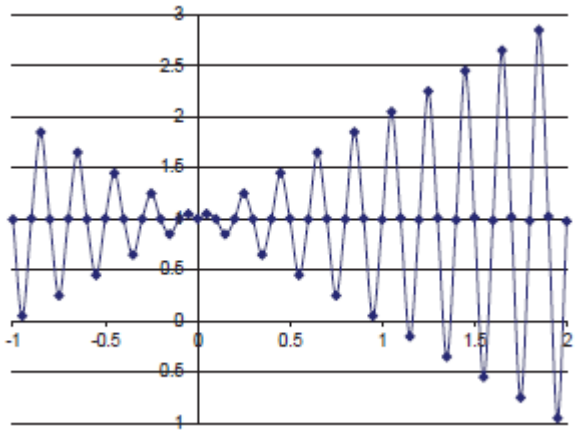
- **Εύρεση Μεγίστου Συνάρτησης μιας Μεταβλητής**
- f(x)=x-sin(10π-x)+1, ορισμένη στο διάστημα [-1, 2]
- Η αντιμετώπιση του προβλήματος αυτού αναλυτικά (δηλαδή μηδενισμός πρώτης παραγώγου, κλπ) είναι αρκετά δύσκολη…
- Το πρόβλημα μπορεί να αντιμετωπισθεί με ΓΑ!
- Χρήση **δυαδικής αναπαράστασης** 1
  - Έστω ότι θέλουμε ακρίβεια 6 δεκαδικών ψηφίων.
  - Αυτά μας δίνουν 10^6 συνδυασμούς στο διάστημα μιας μονάδας. Για αναπαράσταση του διαστήματος [-1, 2] (=>3 μονάδες) απαιτούνται $3 \cdot 10^6$ αριθμοί.

Παράδειγμα 1

- **Εύρεση Μεγίστου Συνάρτησης μιας Μεταβλητής**
- f(x)=x-sin(10π-x)+1, ορισμένη στο διάστημα [-1, 2]
  - Επειδή $2^{21} < 3 \cdot 10^6 < 2^{22}$, χρειάζονται 22 δυαδικά ψηφία.
  - Η αντιστοίχηση μεταξύ δυαδικών $<b_{21}, b_{20}, ..., b_0>$ και δεκαδικών στο [-1, 2] γίνεται ως εξής:
    - $x = -1 + x' \cdot (3/(2^{22}-1))$
    - όπου x η παράμετρος της f(x) και x' ο δεκαδικός που αντιστοιχεί στη δυαδική αναπαράσταση.

Παράδειγμα 1

- **Εύρεση Μεγίστου Συνάρτησης μιας Μεταβλητής**
- f(x)=x-sin(10π-x)+1, ορισμένη στο διάστημα [-1, 2]
  - Η συνάρτηση καταλληλότητας είναι η ίδια η f(x), ενώ ο αρχικός πληθυσμός δημιουργείται τυχαία. Εφαρμόζοντας κλασικές τεχνικές διασταύρωσης και μετάλλαξης υπολογίζεται ότι το μέγιστο βρίσκεται "περίπου" στη θέση x=1.85 με f(1.85)=2.85.



Παράδειγμα 2

- Το πρόβλημα του **Πλανόδιου Πωλητή** (TSP)
- Εύρεση της σειράς με την οποία ένας πωλητής πρέπει να περάσει από όλες τις πόλεις ενός συνόλου πόλεων και να επιστρέψει στην αρχική, ώστε να έχει το μικρότερο δυνατό κόστος, όπως αυτό εκφράζεται κάθε φορά (απόσταση, χρόνος, κλπ).
- **Αναπαράσταση** με **διανύσματα ακεραίων αριθμών** μήκους ίσου με το πλήθος των πόλεων. Οι αριθμοί δεν επαναλαμβάνονται: 1
- $v=<i_1 i_2 \ldots i_n>$, όπου $i_1, i_2, \ldots, i_n \in 1..n$ και $i_j \neq i_k$ για $j \neq k$

Παράδειγμα 2

- Το πρόβλημα του **Πλανόδιου Πωλητή**
- **Δημιουργία αρχικού πληθυσμού** με τυχαίο τρόπο.
- Απλή **συνάρτηση καταλληλότητας**: 2
  - π.χ. **άθροισμα αποστάσεων** μεταξύ πόλεων
- **Αναπαραγωγή**: Πρέπει να προκύπτουν **έγκυρα** (βάσει προβλήματος) **χρωμοσώματα**!
  - π.χ. να μην εμφανίζεται μία πόλη δύο φορές!

- (Λεπτομέρειες για το μηχανισμό αναπαραγωγής θα δούμε παρακάτω…)

Γ. Αρχικοποίηση Πληθυσμού

- Παράγονται "λύσεις" (όχι απαραίτητα καλές – ζητούμενο είναι να συμφωνούν με την αναπαράσταση που έχει επιλεγεί!).
  - Συνήθως είναι τυχαίες τιμές που παράγονται από μια γεννήτρια τυχαίων αριθμών.
- **Προγραμματιστικά**, ο πληθυσμός μιας γενιάς είναι **μια δομή δεδομένων πίνακα** (ή συνδεδεμένη λίστα). Το μέγεθός της καθορίζεται από το μέγεθος του πληθυσμού με τον οποίο θέλουμε να δουλέψουμε.
- Το **μέγεθος** του **πληθυσμού** πρέπει να είναι **επαρκώς μεγάλο** αλλά όχι πολύ μεγάλο! Μπορούμε να το θεωρήσουμε όπως τη διαδικασία δειγματοληψίας από ένα σύνολο.
  - Αν είναι πολύ μικρό, η αναζήτηση θα καθυστερήσει και ίσως δεν δώσει και καλό αποτέλεσμα.
  - Αν είναι πολύ μεγάλο, θα υπάρχει υπολογιστική επιβάρυνση (χρόνος, μνήμη).
- **Τυπικά μεγέθη** είναι μερικές **εκατοντάδες** (χωρίς να είναι δεσμευτικό).
  - Ζητούμενο (αλλά όχι πάντα εύκολο να γίνει!) είναι να υπάρχει επαρκής εκπροσώπηση λύσεων από όλες τις περιοχές του χώρου των λύσεων!

Δ. Μηχανισμός Επιλογής Γονέων

- Θέλουμε τα **περισσότερο ποιοτικά άτομα** να έχουν μεγαλύτερη πιθανότητα επιβίωσης (άρα και αναπαραγωγής) από τα λιγότερο ποιοτικά.
- Αυτό αντιγράφει την Θεωρία της Εξέλιξης! Θεωρούμε ότι η ποιότητά τους οφείλεται στο "γενετικό" τους υλικό – άρα θέλουμε στοιχεία αυτού του υλικού να περάσουν στις επόμενες γενιές με **μεγαλύτερη συχνότητα**.
- Τα αδύναμα άτομα δεν "καταδικάζονται" – απλά συμμετέχουν με **μικρότερη συχνότητα**.
- Στη φύση, τέτοια άτομα ίσως να έχουν κάποια χρήσιμα για τις επόμενες γενιές χαρακτηριστικά!
- Υπάρχουν πολλοί μηχανισμοί στη βιβλιογραφία. Αρκετά διαδεδομένοι είναι αυτοί που βασίζονται στην έννοια του **Τροχού Επιλογής**. Θα μελετήσουμε έναν αντιπροσωπευτικό:

Επιλογή με Ρουλέτα

- Έστω ότι θέλουμε το μέγιστο της συνάρτησης

  $y = -(1/4)\, x^2 + 2x + 5$, στο [0, 10]

  - Άρα η συνάρτηση καταλληλότητας είναι η ίδια η παραπάνω συνάρτηση.
- Έστω ότι επιλέγουμε **αναπαράσταση** 10-bits.
- Τα 10-bits επιτρέπουν τιμές μεταξύ 0 και 2^{10} , δηλαδή μεταξύ 0 και 1024. Μπορούμε εύκολα να αναγάγουμε αυτές τις τιμές σε κάποιο άλλο επιθυμητό διάστημα τιμών, εδώ στο [0,10].
- Ο ακόλουθος πίνακας περιέχει τα μεγέθη που απαιτεί η επιλογή με ρουλέτα.

Επιλογή με Ρουλέτα

- Έστω ότι θέλουμε το μέγιστο της συνάρτησης

  $y = -(1/4)\,x^2 + 2x + 5$, στο [0, 10]

α/α	χρωμόσωμα σε δυαδική	χρωμόσωμα σε βάση 10	x	fitness f(x)	% στο σύνολο	αθροιστική πιθανότητα
1	0001101011	107	1.05	6.82	31%	31%
2	1111011000	984	9.62	1.10	5%	36%
3	0100000101	261	2.55	8.47	38%	74%
4	1110100000	928	9.07	2.57	12%	86%
5	1110001011	907	8.87	3.07	14%	100%
			Σύνολο:	22.05	100%	

- Ο πίνακας περιέχει:
  - 5 τυχαία άτομα (χρωμοσώματα)
  - την ισοδύναμη δεκαδική τους τιμή
  - αναγωγή αυτής στο διάστημα [0, 10]
  - την τιμή της συνάρτησης καταλληλότητας
  - το ποσοστό που αντιπροσωπεύει αυτή η τιμή στο άθροισμα (σύνολο) των τιμών καταλληλότητας (στο 22.05, κάτω), και τέλος…

Επιλογή με Ρουλέτα

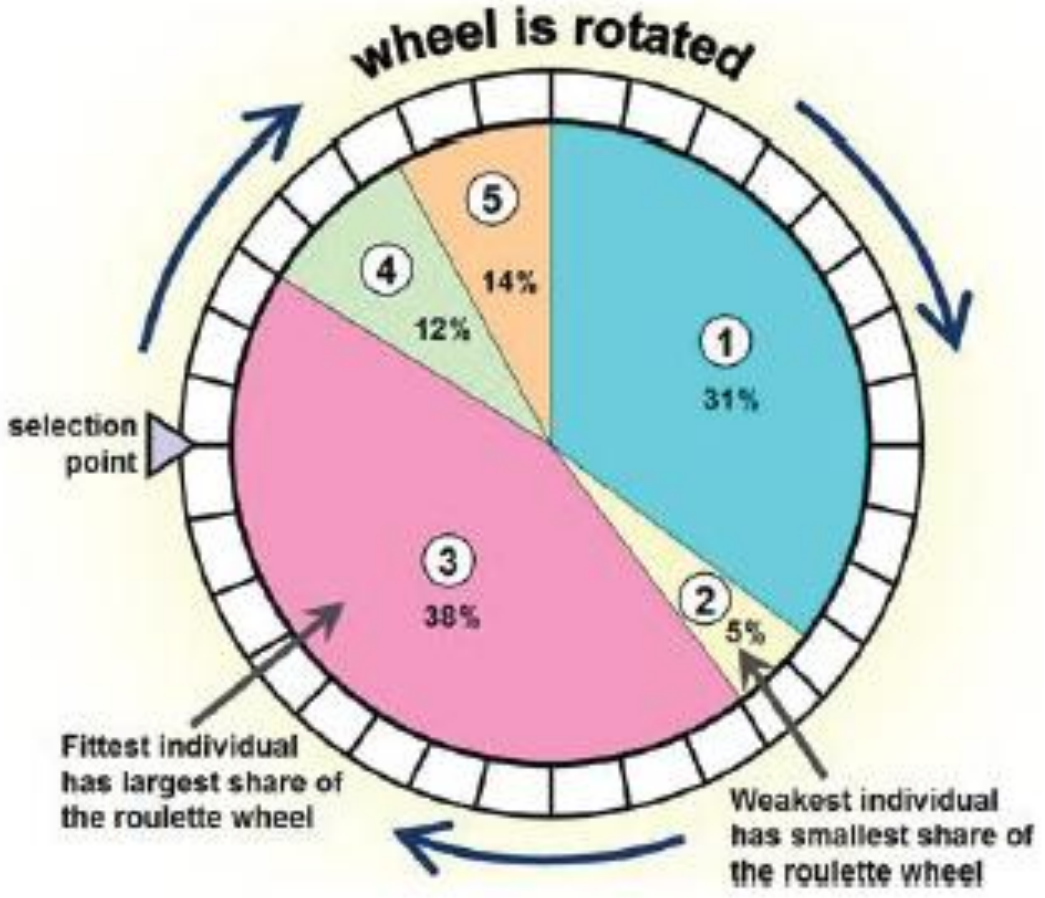
- Έστω ότι θέλουμε το μέγιστο της συνάρτησης

  $y = -(1/4)\,x^2 + 2x + 5$, στο [0, 10]

α/α	χρωμόσωμα σε δυαδική	χρωμόσωμα σε βάση 10	x	fitness f(x)	% στο σύνολο	αθροιστική πιθανότητα
1	0001101011	107	1.05	6.82	31%	31%
2	1111011000	984	9.62	1.10	5%	36%
3	0100000101	261	2.55	8.47	38%	74%
4	1110100000	928	9.07	2.57	12%	86%
5	1110001011	907	8.87	3.07	14%	100%
			Σύνολο:	22.05	100%	

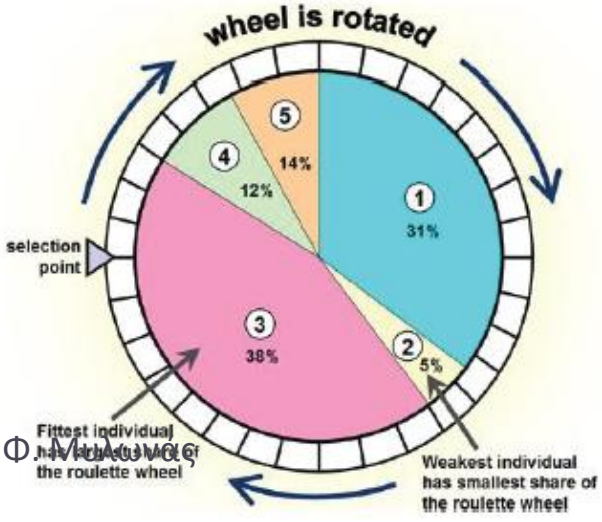
- Ο πίνακας περιέχει:
  - την αθροιστική πιθανότητα (τελευταία στήλη)
    - σε κάθε κελί αθροίζουμε τις μέχρι εκεί πιθανότητες

Επιλογή με Ρουλέτα



Επιλογή με Ρουλέτα

- Η προτελευταία στήλη του πίνακα καθορίζει το μέγεθος των τομέων πάνω στον τροχό.
- Ο τροχός γυρίζει τόσες φορές όσα και τα άτομα που θέλουμε να επιλέξουμε.
- Παράγουμε έναν τυχαίο αριθμό μεταξύ 0-100, έστω Κ. Με βάση αυτόν, **επιλέγουμε το χρωμόσωμα με την μεγαλύτερη αθροιστική πιθανότητα που δεν υπερβαίνει το Κ.** Παραδείγματα: Κ=10 → #1, Κ=31 → #1, Κ=80 → #3, Κ=99 → #5
- Η μέθοδος μεροληπτεί υπέρ των "ποιοτικότερων" ατόμων. Αν το κάνει πολύ έντονα, υπάρχουν τεχνικές αναπροσαρμογής των τιμών για αποφυγή πρόωρης σύγκλισης (τοπικό μέγιστο).

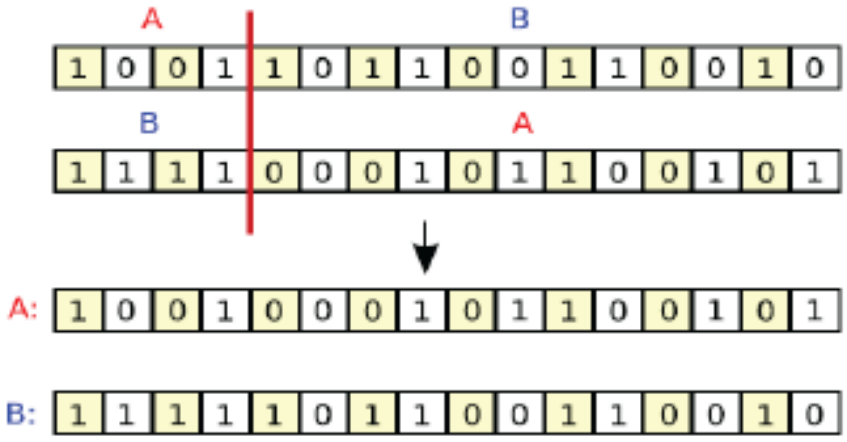


Ε. Αναπαραγωγή / Ανασυνδυασμός

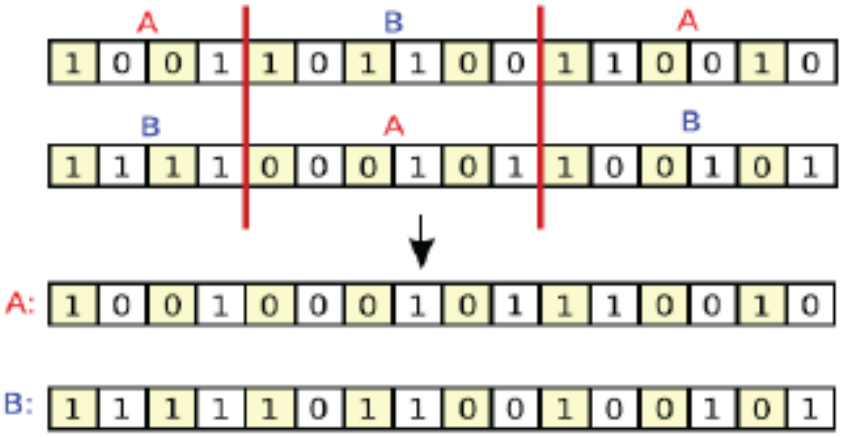
- Έχοντας δημιουργήσει τον πληθυσμό που θα συμμετάσχει στην αναπαραγωγική διαδικασία, **επιλέγουμε** από αυτόν **τυχαία ζευγάρια** και εφαρμόζουμε **τεχνικές ανασυνδυασμού** (**αναπαραγωγής**).
- Υπάρχουν πάρα πολλές τεχνικές. Το τι θα επιλέξουμε το καθορίζει κύρια ο τρόπος περιγραφής που έχουμε υιοθετήσει αρχικά. Θα αναφέρουμε χαρακτηριστικές περιπτώσεις για τις αναπαραστάσεις:
  - δυαδική και
  - permutation

Αναπαραγωγή σε Δυαδική αναπαράσταση

❑ **Single Point Crossover**



❑ **N Point Crossover**



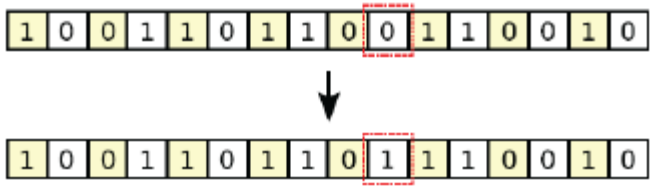
- A και B (κάτω) είναι οι δύο απόγονοι. Επιπλέον στους γονείς (πάνω) είναι σημειωμένο ποιο τμήμα πάει στον απόγονο A και ποιο στον B. Τα τμήματα οριοθετούνται με μια γραμμή σε τυχαία θέση (η κόκκινη γραμμή).

Partially Mapped Crossover - Permutation

- Ας τη δούμε με παράδειγμα στο πρόβλημα του πλανόδιου πωλητή (TSP):
- Έστω ότι οι πόλεις είναι 9. Υιοθετήσαμε νωρίτερα την περιγραφή με μια αλληλουχία των αριθμών 1 ως 9, χωρίς να επαναλαμβάνονται οι αριθμοί.
- Έστω p1=(1 2 3 | 4 5 6 7 | 8 9) και p2=(4 5 2 | 1 8 7 6 | 9 3) δύο τυχαίες λύσεις.
  - τα σύμβολα | μπήκαν τυχαία – ορίζουν τα τμήματα που θα χρησιμοποιηθούν
- Δημιουργείται η αρχική μορφή των απογόνων κρατώντας το κεντρικό τμήμα γονέων
  - o_1=(x x x | **4 5 6 7** | x x) και o_2=(x x x | **1 8 7 6** | x x)
- Για να συμπληρωθούν τα υπόλοιπα στοιχεία του o_1, θα χρησιμοποιηθεί το χρωμόσωμα του γονέα p_2 αναδιαταγμένο.
  - p_2=(4 5 2 | 1 8 7 6 | **9 3**) αναδιάταξη: **9 3** 4 5 2 1 8 7 6
- Επιπλέον αφαιρούνται οι πόλεις (ψηφία) που υπάρχουν ήδη στο o_1=(xxx|4567|xx)
  - δηλαδή από το 9 3 4 5 2 1 8 7 6 απομένει το 9 3 2 1 8
  - με τα ψηφία 9 3 2 1 8 συμπληρώνεται η λύση o1 αρχίζοντας από το τέλος (με την ίδια σειρά δηλαδή που έγινε αναδιάταξη στον γονέα p2).
- Οπότε γίνεται: o1=(218|4567|93)
- Όμοια προκύπτει ότι o2=(345|1876|92)

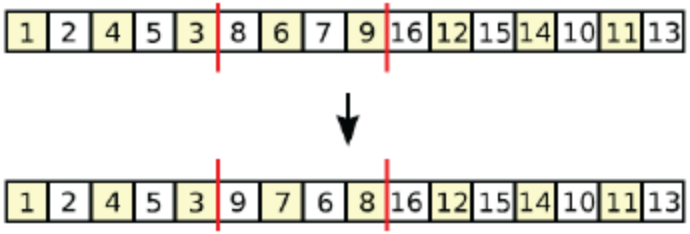
Μετάλλαξη (Mutation)

- **Σε Δυαδική Αναπαράσταση**: με χαμηλή συχνότητα, επιλέγουμε ένα τυχαίο bit σε τυχαίο άτομο και του αλλάζουμε τιμή.
- Ο ρυθμός μετάλλαξης υλοποιείται εύκολα. Αν π.χ. θέλουμε 1 στις 1000 φορές, παράγουμε έναν ακέραιο στο διάστημα (1,1000). Αν είναι π.χ. 0 τότε κάνουμε τη μετάλλαξη, αλλιώς όχι.

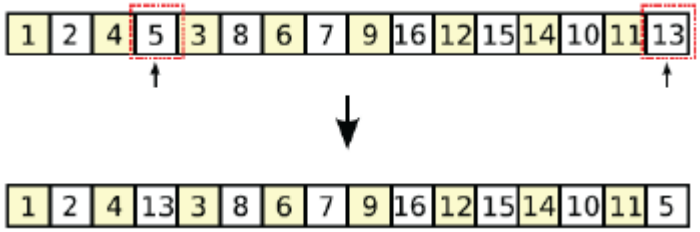


Μετάλλαξη (Mutation)

- **Σε Αναπαράσταση Permutation** (Μετάθεσης):
- **Inversion/Αναστροφή**: μια τυχαία αλληλουχία αριθμών, μικρού μήκους, αναστρέφεται.



- **Swap**: δύο τυχαία επιλεγμένοι αριθμοί ενός χρωμοσώματος, ανταλλάσουν θέση.



ΣΤ. Ορισμός του Πληθυσμού της Επόμενης Γενιάς

- Ζητούμενο: να καθορίσουμε ποια άτομα (από την τρέχουσα γενιά και τους απογόνους) θα αποτελέσουν την επόμενη γενιά.
- Επιλέγουμε:
  - **μ** άτομα από το σύνολο της τρέχουσας γενιάς
  - **λ** απογόνους (τα άτομα που προέκυψαν από την αναπαραγωγή)
- Επιπλέον μπορούμε να κρατούμε το καλύτερο "άτομο".
  - **Ελιτισμός** – elitism: Η καλύτερη λύση (στο σύνολο αρχικού πληθυσμού και παιδιών) μεταφέρεται απευθείας στην επόμενη γενιά ώστε κατ' ελάχιστο να μη χειροτερεύει η καλύτερη λύση (άτομο) που έχουμε.
- 2 διαδεδομένοι τρόποι είναι **με βάση την ηλικία** και **με βάση την ποιότητα**.

Επιλογή ως προς την **Ηλικία**

- Μέρος της τρέχουσας γενιάς αντικαθίσταται από απογόνους. Δύο τυπικοί τρόποι:
  - Τα λ χειρότερα άτομα της τρέχουσας γενιάς αντικαθίστανται από λ απόγονους.
  - Τυχαία, λ άτομα της τρέχουσας γενιάς αντικαθίστανται από λ απόγονους

Επιλογή ως προς την **Ποιότητα**

- Κριτήριο Επιλογής είναι η τιμή ποιότητας. Δύο τυπικοί τρόποι:
- **Tournament**: Από το σύνολο ατόμων (απόγονοι και τρέχουσα γενιά), επιλέγεται ένα μικρό υποσύνολο k ατόμων και το άτομο με τη χειρότερη ποιότητα απομακρύνεται. Η διαδικασία επαναλαμβάνεται μέχρι να απομείνει το επιθυμητό πλήθος ατόμων.
- **GENITOR**: Από το σύνολο ατόμων (απόγονοι και τρέχουσα γενιά), απομακρύνονται τα χειρότερα άτομα ώστε να μείνει το επιθυμητό πλήθος.
  - Είναι λίγο επικίνδυνη μέθοδος γιατί μπορεί να οδηγήσει τον ΓΑ σε πρόωρη σύγκλιση σε τοπικό ακρότατο (δεν θα τον αφήσει να διερευνήσει επαρκώς τον χώρο αναζήτησης (δηλ. τον χώρο των λύσεων)).

Συνθήκη Τερματισμού ΓΑ

- Ο ΓΑ εκτελείται επαναληπτικά έως ότου ικανοποιηθεί η συνθήκη ή οι συνθήκες τερματισμού που έχουν οριστεί. Συνήθως η συνθήκη τερματισμού ορίζεται από το πρόβλημα, ωστόσο οι περισσότερο συνηθισμένες είναι οι ακόλουθες:
  - Μέγιστο **πλήθος επαναλήψεων**
  - Μέγιστος **χρόνος εκτέλεσης** του ΓΑ
  - Η **μη βελτίωση** της ποιότητας του καλύτερου ατόμου για ένα **προκαθορισμένο πλήθος γενεών**
  - Η **εύρεση** της **βέλτιστης** ή μιας **αποδεκτής λύσης**

Σύγχρονες Εφαρμογές ΓΑ 1/2

- Εύρεση **μέγιστης τιμής αριθμητικών συναρτήσεων**
  - Η εύρεση του μέγιστου μιας συνάρτησης δεν είναι καθόλου εύκολη υπόθεση για συναρτήσεις πολλών μεταβλητών, οι οποίες εμφανίζουν ασυνέχειες, θόρυβο, κλπ.
  - Το πλεονέκτημα που εμφανίζει η εφαρμογή τους σε αυτά τα προβλήματα είναι ότι η συνάρτηση καταλληλότητας είναι η ίδια η συνάρτηση του προβλήματος!
- **Επεξεργασία Εικόνων**
  - Αναγνώριση προτύπων, όπως ακμές, επιφάνειες, αντικείμενα, σε ψηφιοποιημένες εικόνες.
- **Συνδυαστική βελτιστοποίηση**
  - Το κλασσικό πρόβλημα κατανομής πόρων σε δραστηριότητες, με σκοπό τη μεγιστοποίηση του οφέλους ή την ελάττωση του κόστους.
  - Ο έλεγχος όλων των υποψήφιων λύσεων να είναι αδύνατος (συνδυαστική έκρηξη)!
  - Γνωστά προβλήματα αυτής της κατηγορίας: του πλανόδιου πωλητή (TSP), η αποθήκευση κιβωτίων, σχεδίαση VLSI κυκλωμάτων, καταμερισμός εργασιών, ωρολόγιο πρόγραμμα, βελτιστοποιημένη κοπή υλικών για ελαχιστοποίηση απωλειών υλικού (υφάσματα, ελάσματα, ξύλα), τοποθέτηση αναμεταδοτών/κεραιών, κτλ

Σύγχρονες Εφαρμογές ΓΑ 2/2

- Σχεδίαση
  - Σχεδίαση κατασκευών και εξαρτημάτων, με ζητούμενο τόσο την εύρεση μιας λύσης, όσο και τη βελτιστοποίησή της ώστε να πληροί κάποιες ιδιότητες.
  - Σχεδίαση μορίων με επιθυμητές ιδιότητες (για φαρμακευτικές ουσίες, κτλ).
  - Οι αλγόριθμοι μπορούν να δοκιμάσουν συνδυασμούς και ιδέες που ο ανθρώπινος νους δε θα δοκίμαζε ποτέ, δίνοντας ενίοτε πρωτότυπα αποτελέσματα.

Σύνοψη ΓΑ

- Με **εξαίρεση** α) την **αναπαράσταση** και β) την **συνάρτηση καταλληλότητας** που εξαρτώνται από το υπό μελέτη πρόβλημα, τα υπόλοιπα βήματα στους ΓΑ είναι **τυποποιημένα**.
- Οι ΓΑ περιέχουν **αρκετές παραμέτρους** και απαιτούν **αρκετό πειραματισμό** για να βρεθούν οι κατάλληλες τιμές σε αυτές (π.χ. μέγεθος πληθυσμού γενιάς).
- Οι ΓΑ εμπεριέχουν **έντονο στοιχείο τυχαιότητας** (randomness) και **διαδοχικά run ενός αλγορίθμου δεν εγγυώνται παρόμοια αποτελέσματα**!
- Παρόλα αυτά οι ΓΑ έχουν επιδείξει εξαιρετική ικανότητα στο να **μοντελοποιήσουν** φαινόμενα και συμπεριφορές με έντονο συνδυαστικό χαρακτήρα (πολλές παράμετροι, κτλ).

Νευρωνικά Δίκτυα - Εισαγωγή

- Είναι μια ιδιαίτερη προσέγγιση στη δημιουργία συστημάτων με νοημοσύνη.
  - **Δεν** αναπαριστούν ρητά τη γνώση (π.χ. με κανόνες if-then) όπως τα συστήματα γνώσης.
  - **Δεν** υιοθετούν ειδικά σχεδιασμένους αλγόριθμους αναζήτησης.
  - **Δεν** εξελίσσουν καλές λύσεις όπως οι Γενετικοί Αλγόριθμοι
  - Βασίζονται σε **βιολογικά πρότυπα:**
    - προσομοίωση μικρής κλίμακας του τρόπου λειτουργίας του ανθρώπινου εγκεφάλου.

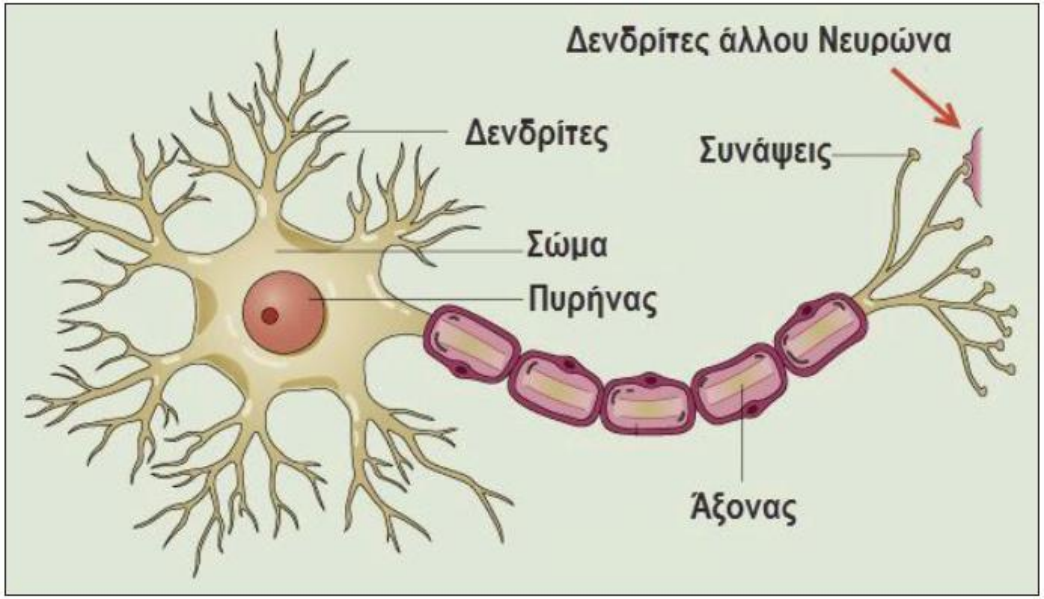
Νευρωνικά Δίκτυα - Εισαγωγή

- Έχουν ικανότητα μάθησης κύρια μέσω καταγεγραμμένων παρατηρήσεων και μπορεί να συνεισφέρουν στην ευφυΐα μιας τεχνητής οντότητα (π.χ. ενός λογισμικού πράκτορα / software agent), με κάποιον από τους ακόλουθους τρόπους:
  - **πρόβλεψη** (π.χ. βραχυπρόθεσμη πρόβλεψη ισοτιμιών νομισμάτων ή τιμών μετοχών)
  - **ταξινόμηση** (π.χ. κατηγοριοποίηση ιατρικών εικόνων, στόχων, κτλ)
  - **αναγνώριση** (π.χ. προσώπου σε συστήματα ασφάλειας)
  - **αποτίμηση** (π.χ. παρακολούθηση στόχων σε οπλικά συστήματα με αυτονομία)

- Πώς θα μπορούσαμε να φτιάξουμε τεχνητές δομές που να λειτουργούν όπως ο εγκέφαλος ?
- Πώς λειτουργεί ο εγκέφαλος ?

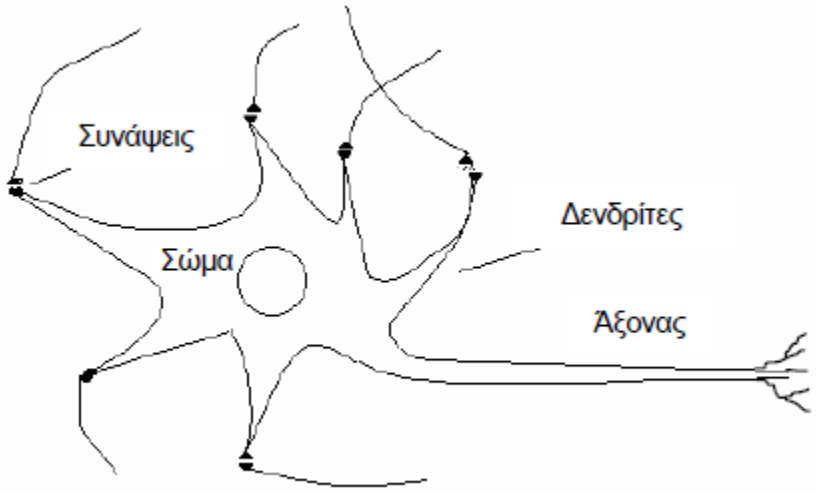
Βιολογικός νευρώνας

- Τα ηλεκτρικά **σήματα** (προερχόμενα από άλλους νευρώνες) που εισέρχονται στο σώμα μέσω των δενδριτών, **συνδυάζονται** και αν το αποτέλεσμα ξεπερνά κάποιο όριο (κατώφλι), παράγεται κάποιο **σήμα εξόδου** το οποίο **διαδίδεται** μέσω του άξονα προς άλλους νευρώνες.



Βιολογικός νευρώνας

- Μάθηση και μνήμη υλοποιούνται με μεταβολές στην αγωγιμότητα των συνάψεων.
  - η αγωγιμότητα μεταβάλλεται με χημικές διεργασίες

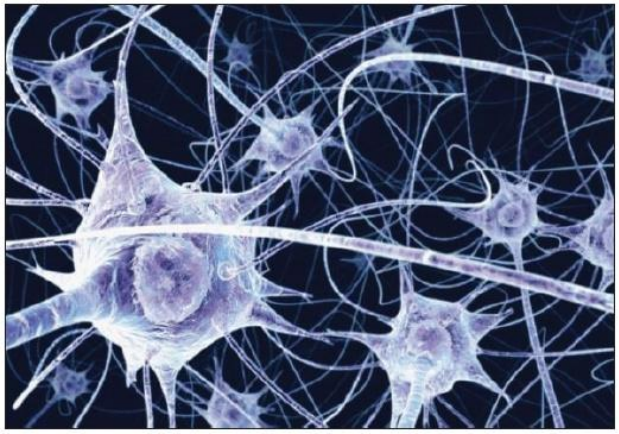


Φυσικά Νευρωνικά Δίκτυα

- Ανθρώπινος εγκέφαλος:
  - περίπου 100 δισ. νευρώνες
  - κάθε νευρώνας συνδέεται κατά μέσο όρο με 1000 άλλους νευρώνες
  - => περίπου 100 τρισ. συνάψεις!
- Η αντιγραφή ομοιώματος είναι εφικτή μόνο σε περιορισμένη κλίμακα
  - μερικές χιλιάδες ή δεκάδες χιλιάδες νευρώνες – συνήθως πολύ λιγότεροι
- Χρόνος απόκρισης των βιολογικών νευρώνων:
  - της τάξης των msec (0.001 second)
  - ...όμως ο εγκέφαλος λαμβάνει πολύπλοκες αποφάσεις, εκπληκτικά γρήγορα.

Φυσικά Νευρωνικά Δίκτυα

- Ανθρώπινος εγκέφαλος:
  - περίπου 100 δισ. νευρώνες
  - κάθε νευρώνας συνδέεται κατά μέσο όρο με 1000 άλλους νευρώνες
  - => περίπου 100 τρισ. συνάψεις!
- Η υπολογιστική ικανότητα του εγκεφάλου και η πληροφορία που περιέχει είναι διαμοιρασμένα σε όλο του τον όγκο
  - ο εγκέφαλος είναι ένα **παράλληλο** και **κατανεμημένο** υπολογιστικό σύστημα!



Ιστορικά στοιχεία 1/5

- **1943**, McCulloch (νευροφυσιολόγος) και Pitts (μαθηματικός): πρώτο μοντέλο ΤΝΔ στην προσπάθεια εξήγησης της μνήμης
  - ο νευρώνας έχει **πολλές εισόδους** αλλά **μία έξοδο**
  - οι **έξοδοι δεν ενώνονται** – πρέπει όμως **να καταλήγουν σε άλλο νευρώνα**, πιθανώς και στον ίδιο νευρώνα
  - ο νευρώνας μπορεί να είναι σε **δύο καταστάσεις**: **πυροδοτεί** (στέλνει ηλεκτρικό παλμό) ή βρίσκεται **σε ηρεμία**
  - οι λειτουργίες αυτές γίνονται σε διακριτό χρόνο, δηλ. το σύστημα δρα συγχρονισμένα
  - η **κατάσταση σε χρόνο t+1 εξαρτάται** από την **κατάσταση σε χρόνο t και τις εισόδους** που δέχεται εκείνη τη στιγμή
  - Ερμηνεία: η μνήμη υλοποιείται με κλειστές διαδρομές ηλεκτρικού σήματος που ελέγχονται από διεγερτικούς και ανασταλτικούς μηχανισμούς

Ιστορικά στοιχεία 2/5

- **1949**, Hebb: κανόνας μάθησης Hebb
  - κάθε φορά που το σύστημα διεγείρεται με "σήμα εισόδου" και χρησιμοποιεί τις συνδέσεις μεταξύ των νευρώνων, αυτές ενισχύονται και το δίκτυο πλησιάζει περισσότερο στο να μάθει το "σήμα" εισόδου
- **Δεκαετία '50**: John von Newman
  - ανέφερε τις εργασίες των McCulloch & Pitts ως παραδείγματα υπολογιστικών μηχανών
- **1957**, Rosenblatt: έφτιαξε το πρώτο δίκτυο (σε hardware) – perceptron (αισθητήρας)
  - είχε μόνο είσοδο και έξοδο και μπορούσε να κάνει διάφορες διεργασίες (βλ. συνέχεια)
  - αρχικά δημιούργησε μεγάλο ενθουσιασμό
- **1959**, Widrow και Hoff: ανέπτυξαν δύο νέα μοντέλα (Adaline, Madaline) που χρησιμοποιήθηκαν με επιτυχία σε πρακτικές εφαρμογές (ως φίλτρα σε τηλεφωνικά δίκτυα)

Ιστορικά στοιχεία 3/5

- **1969**, Minsky και Papert: απέδειξαν με αναλυτικά μαθηματικά ότι υπάρχουν συγκεκριμένοι περιορισμοί στο τι μπορούν να κάνουν τα perceptrons
  - η εργασία αυτή λειτούργησε ως "ταφόπλακα" για τα ΤΝΔ
  - το ενδιαφέρον στράφηκε στις συμβολικές και με βάση τη λογική μεθόδους

- Η …Αναγέννηση των ΤΝΔ
- **1982**, Hopfield (βιολόγος): σε 5 σελίδες απέδειξε με μαθηματικό τρόπο:
  - πώς ένα ΤΝΔ μπορεί να χρησιμοποιηθεί ως **αποθηκευτικός χώρος**
  - ότι ένα ΤΝΔ είναι δυνατό να ανακτήσει την αποθηκευμένη πληροφορία **ακόμη κι αν** η **είσοδος διαφέρει ελαφρώς** από αυτά που «γνωρίζει»
- Κατανοήθηκε η απαίτηση για μηχανισμό εκπαίδευσης που να βασίζεται στο σφάλμα που παράγει το δίκτυο στην έξοδό του.

Ιστορικά στοιχεία 4/5

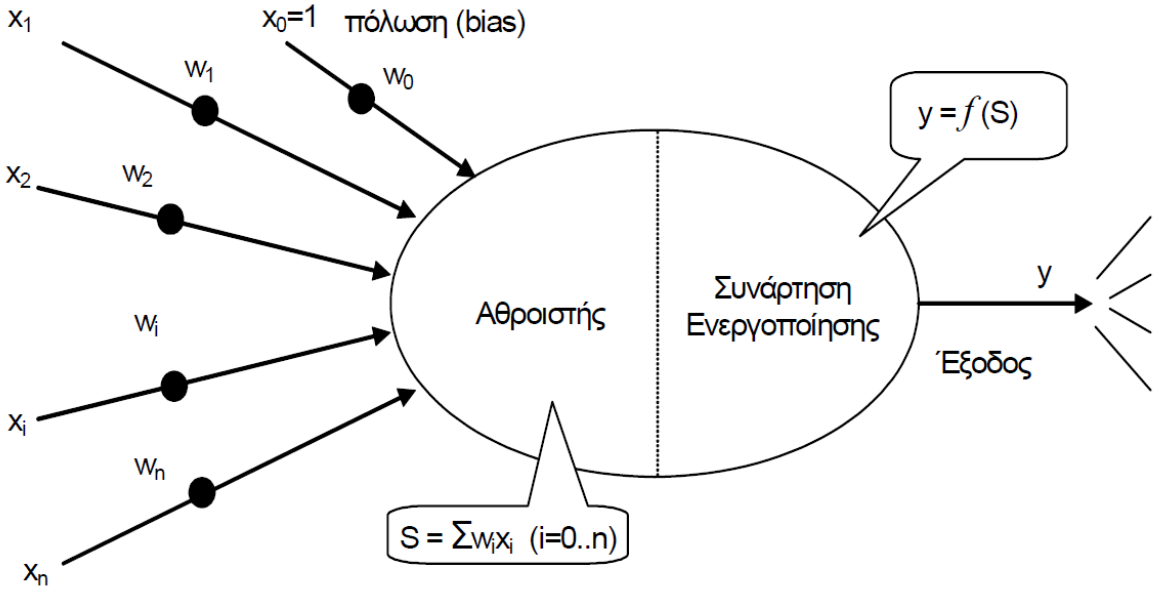
- **1986**, McClelland και Rumelhart: παρουσιάζουν τον τρόπο με τον οποίο ένα ΤΝΔ μπορεί να θεωρηθεί και να χρησιμοποιηθεί ως παράλληλος επεξεργαστής.
  - εισάγουν την έννοια των κρυφών επιπέδων που δεν υπήρχαν στο perceptron
  - προτείνουν την πιο πολυχρησιμοποιημένη (ακόμη και στις μέρες μας) διαδικασία εκπαίδευσης για ΤΝΔ, τη μέθοδο της οπισθοδιάδοσης (back-propagation)
    - η μέθοδος αυτή είχε συζητηθεί και παλαιότερα αλλά τώρα διατυπώθηκε ολοκληρωμένα και με αυστηρά μαθηματικό τρόπο

Ιστορικά στοιχεία 5/5

- **Δεκαετία '90**: ραγδαία εξέλιξη
  - ανεξάρτητο επιστημονικό πεδίο
  - τουλάχιστο 10 επιστημονικά περιοδικά αφιερωμένα σε ΤΝΔ
  - ετήσια συνέδρια
  - εμπορικές εφαρμογές (κυρίως από τις ΗΠΑ)

- **Σήμερα**: χαμηλότεροι ρυθμοί εξέλιξης, χωρίς ραγδαία αύξηση των εφαρμογών

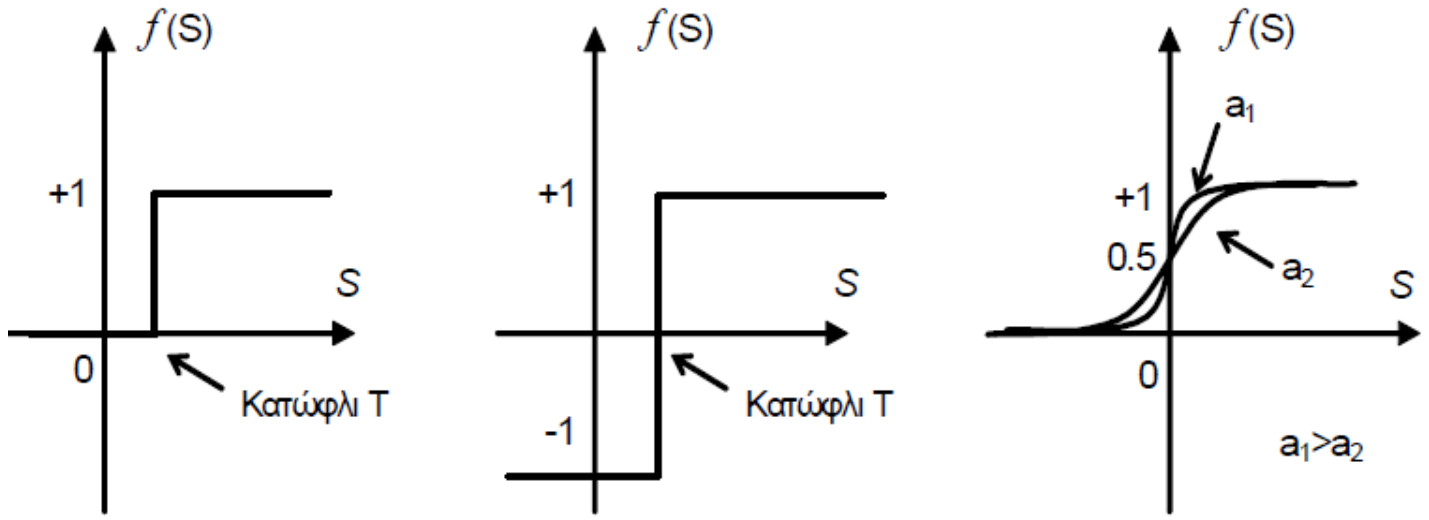
Μοντέλο Τεχνητού Νευρώνα (Artificial Neuron)



- Τα **βάρη w_i** είναι το **ισοδύναμο** των **συνάψεων** του βιολογικού νευρώνα!
- Συνηθέστερη συνάρτηση ενεργοποίησης είναι η **σιγμοειδής συνάρτηση** (βλ. σχήμα)
  - Ονομάζεται έτσι λόγω της γραφικής της παράστασης (πεπλατυσμένο S)
- **Μία έξοδος**: εννοούμε ότι προκύπτει **μια τιμή**
  - οι απολήξεις μπορεί να είναι πολλές!

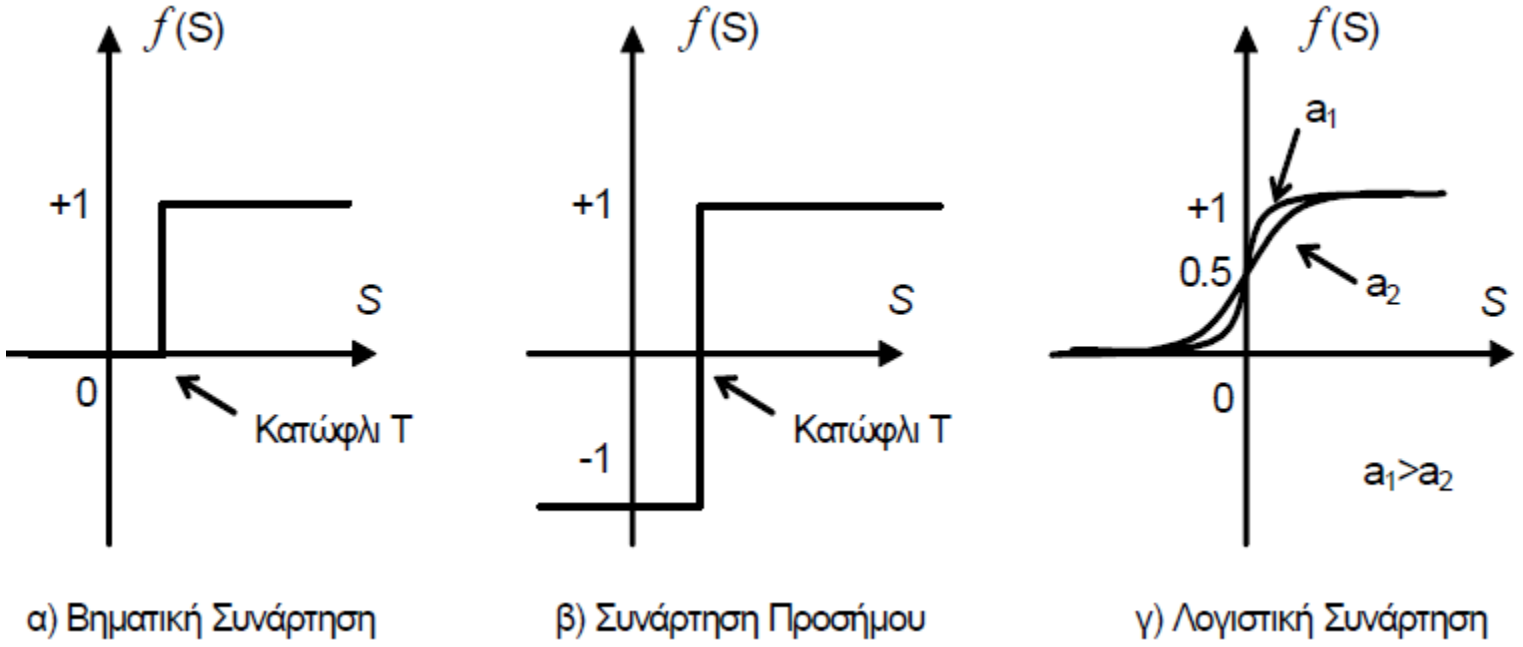
Συναρτήσεις Ενεργοποίησης 1/4

- Βασική απαίτηση: **να είναι μη γραμμική** ώστε να μπορεί να μοντελοποιεί μη γραμμικά φαινόμενα.
- Επίσης πρέπει να είναι **συνεχής** και **παραγωγίσιμη** σε όλο το πεδίο ορισμού της, καθώς το απαιτεί το μαθηματικό μοντέλο πίσω από το μηχανισμό εκπαίδευσης του νευρωνικού δικτύου (θα τον δούμε παρακάτω).



α) Βηματική Συνάρτηση β) Συνάρτηση Προσήμου γ) Λογιστική Συνάρτηση

Συναρτήσεις Ενεργοποίησης 2/4



α) Βηματική Συνάρτηση β) Συνάρτηση Προσήμου γ) Λογιστική Συνάρτηση

- Η **βηματική** και η **προσήμου** χρησιμοποιήθηκαν παλαιότερα σε απλά perceptron.
- Η **λογιστική** (logistic) συνάρτηση είναι μια σιγμοειδής συνάρτηση που έχει τα ζητούμενα χαρακτηριστικά.

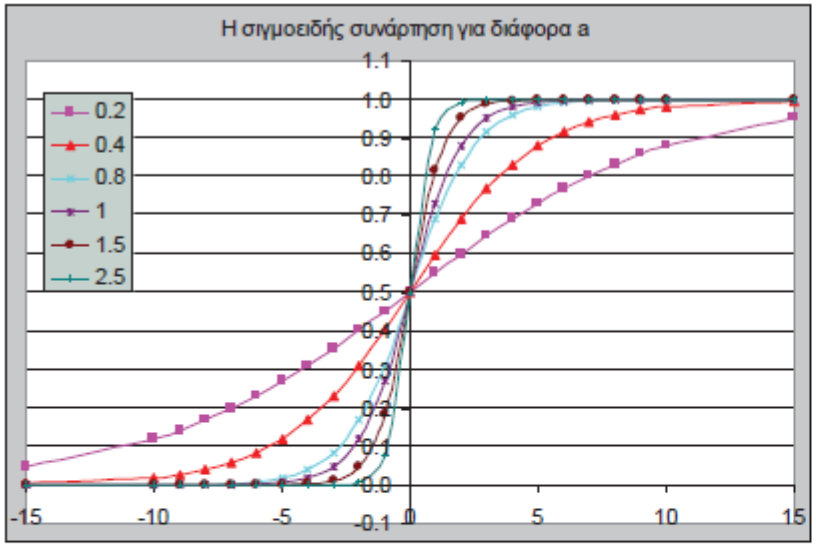
Συναρτήσεις Ενεργοποίησης 3/4

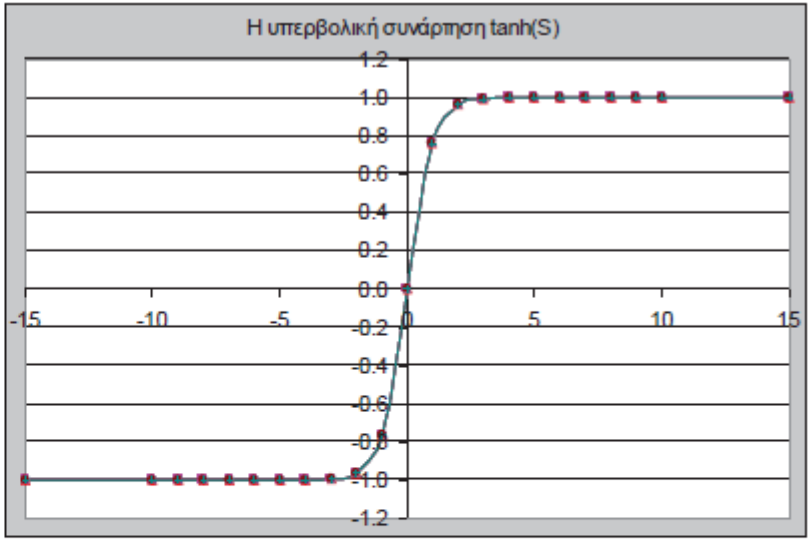
- Η **λογιστική** (logistic) συνάρτηση - μέλος οικογένειας **σιγμοειδών** συναρτήσεων:

$$\Phi(\mathrm{S}) = \frac{1}{1 + e^{-a \cdot \mathrm{S}}}$$

Συναρτήσεις Ενεργοποίησης 4/4

- Γραφική παράσταση των **πιο συχνών σε χρήση** συναρτήσεων ενεργοποίησης σε ΤΝΔ:





λογιστική (logistic) συνάρτηση

$$\Phi(S) = \frac{1}{1+e^{-a \cdot S}}$$

Έξοδος μεταξύ 0 και 1

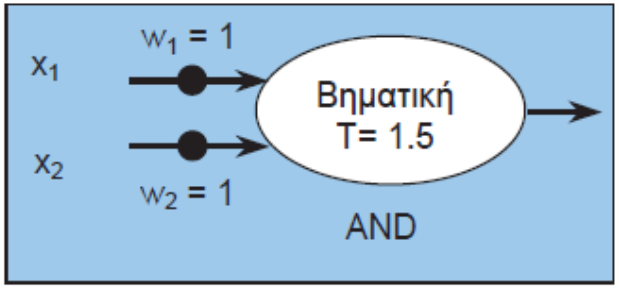
υπερβολική συνάρτηση

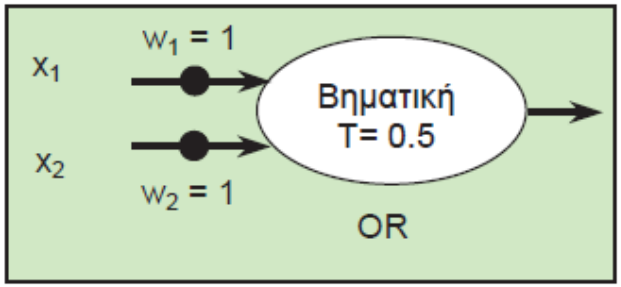
$$\Phi(S) = \tanh(S) = \frac{e^{2 \cdot S} - 1}{e^{2 \cdot S} + 1}$$

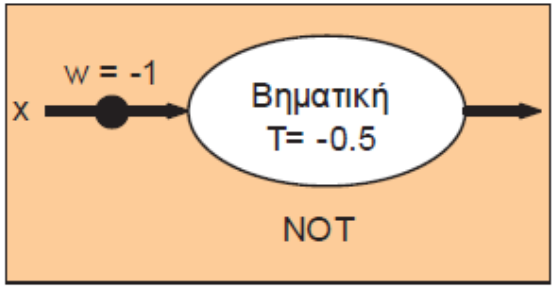
Έξοδος μεταξύ -1 και 1

- Μικρότερες τιμές a στην λογιστική συνάρτηση κάνουν την καμπύλη πιο πεπλατυσμένη.

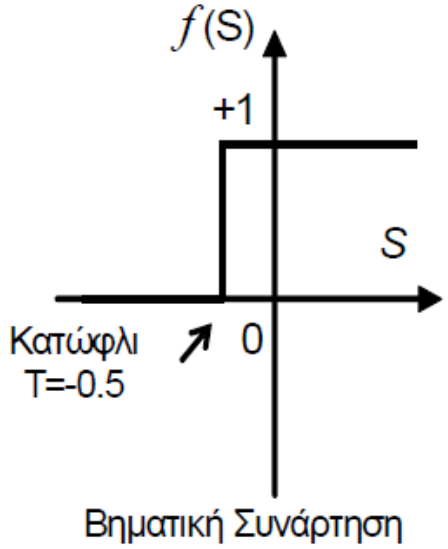
Υλοποίηση **Λογικών** Συναρτήσεων με Τεχνητό Νευρώνα



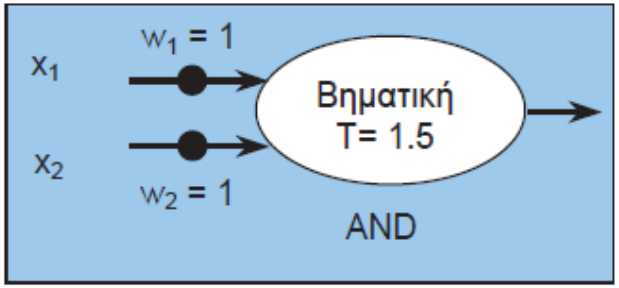


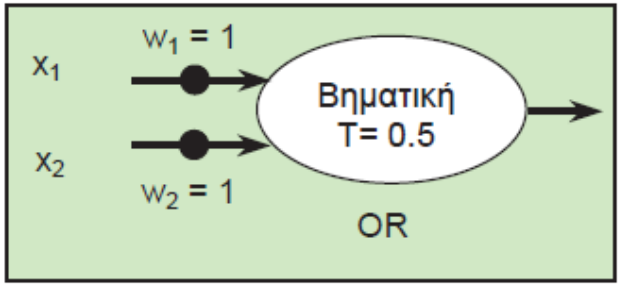


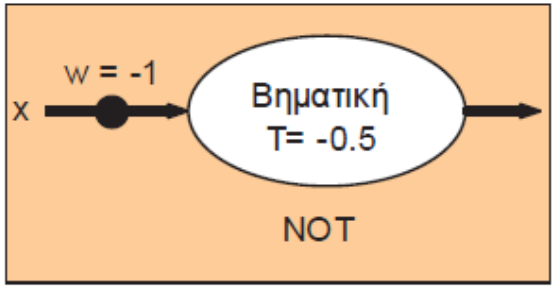
- Στα επόμενα θεωρούμε binary input (0 ή 1)
- Παράδειγμα: υλοποίηση του **NOT**:
  - βηματική συνάρτηση ενεργοποίησης κατώφλι T=-0.5



Υλοποίηση **Λογικών** Συναρτήσεων με Τεχνητό Νευρώνα







- Παράδειγμα: υλοποίηση του **AND**:
  - Όταν αθροιστής βγάλει τιμή πάνω από T=1.5 η έξοδος είναι 1.

x_1	x_2	S	y
1	1	2	1
1	0	1	0
0	1	1	0
0	0	0	0

Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ)

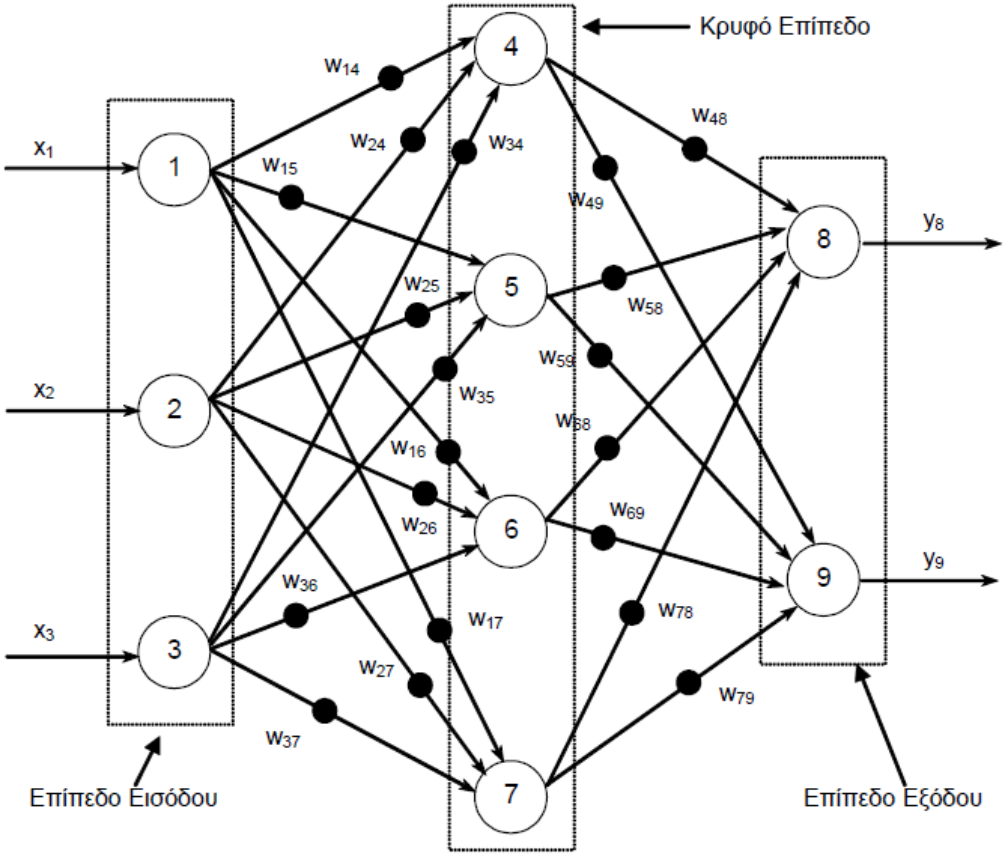
- Συστήματα επεξεργασίας δεδομένων που αποτελούνται από ένα **πλήθος τεχνητών νευρώνων** οργανωμένων σε δομές παρόμοιες με αυτές του ανθρώπινου εγκεφάλου.

- Συντομογραφία για πολυεπίπεδα ΤΝΔ:

  $(p, m_1, m_2, ..., m_q, n)$

Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ)

- Πλήρως συνδεδεμένο ΤΝΔ απλής τροφοδότησης 3-4-2:
- Για το σχήμα: p=3, m_1=4, n=2



Χαρακτηριστικά – Ορολογία

- Οι νευρώνες των διαφόρων στρωμάτων μπορεί να είναι:
  - Πλήρως συνδεδεμένοι (**fully connected**)
  - Μερικώς συνδεδεμένοι (**partially connected**)
- Τα ΤΝΔ χαρακτηρίζονται ως:
  - Δίκτυα με πρόσθια τροφοδότηση (**feedforward**)
  - Δίκτυα με ανατροφοδότηση (**feedback** ή **recurrent**)
- Στην πλειοψηφία των εφαρμογών χρησιμοποιούνται **δίκτυα απλής τροφοδότησης**.

Μάθηση και Ανάκληση

- **Εκπαίδευση** – training (ή **Μάθηση** - learning) είναι η διαδικασία της τροποποίησης της τιμής των βαρών του δικτύου, ώστε **δοθέντος συγκεκριμένου διανύσματος εισόδου να παραχθεί συγκεκριμένο διάνυσμα εξόδου**.
- **Ανάκληση** (recall) είναι η διαδικασία του υπολογισμού ενός διανύσματος εξόδου για **συγκεκριμένο διάνυσμα εισόδου και τιμές βαρών**.
- 3 είδη μάθησης:
  - **Μάθηση με Επίβλεψη** (supervised learning)
  - **Μάθηση χωρίς Επίβλεψη** (unsupervised learning)
  - **Βαθμολογημένη Μάθηση** (graded learning)

Στην πράξη, στις περισσότερες εφαρμογές ΤΝΔ χρησιμοποιείται **μάθηση με επίβλεψη**, για την οποία υπάρχουν αρκετοί αλγόριθμοι.

Μάθηση και Ανάκληση

- **Μάθηση με Επίβλεψη** (supervised learning)
- Στις περισσότερες εφαρμογές ΤΝΔ χρησιμοποιείται μάθηση υπό επίβλεψη
  - ...δηλαδή τροφοδοτούμε το ΤΝΔ με γνωστά δεδομένα εισόδου-εξόδου και ως αποτέλεσμα της εκπαίδευσης **θέλουμε το ΤΝΔ να "μάθει" την άγνωστη σχέση μεταξύ εισόδου-εξόδου**...
  - ...έτσι ώστε δίνοντας στο εκπαιδευμένο ΤΝΔ μία άγνωστη είσοδο, αυτό να μας επιστρέψει την έξοδο (απάντηση) βάσει όσων "έμαθε".

Μάθηση και Ανάκληση

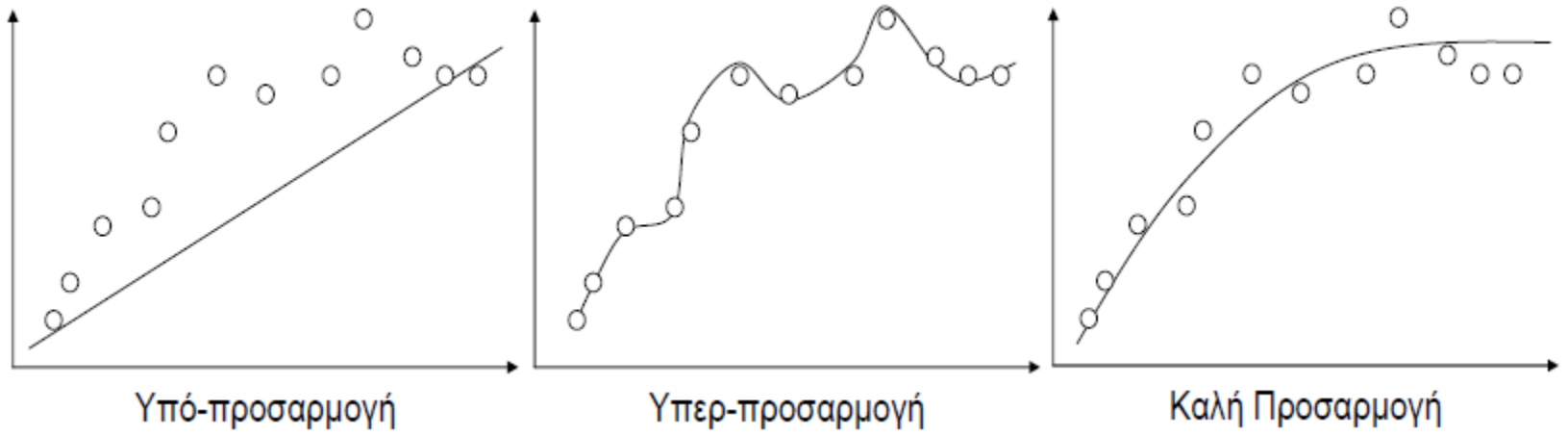
- **Συνολική προσέγγιση:**
- Θεωρούμε ότι τα ζευγάρια input και output εκπαίδευσης, **συσχετίζονται με κάποια άγνωστη σχέση** (που θέλουμε να μάθουμε). Δείχνουμε στο ΤΝΔ γνωστά ζευγάρια input-output και του ζητάμε (μέσω της εκπαίδευσής του) να μοντελοποιήσει αυτή τη σχέση, δηλαδή να είναι σε θέση μετά την εκπαίδευση, όταν του δίνουμε άγνωστα input από το ίδιο πρόβλημα/φαινόμενο, να υπολογίζει σωστό output.

Αλγόριθμοι Μάθησης με Επίβλεψη

- Κανόνας Δέλτα (**Delta rule learning**)
- Αλγόριθμος ανάστροφης μετάδοσης λάθους (**back propagation**)
- Ανταγωνιστική μάθηση (**competitive learning**)
- Τυχαία μάθηση (**random learning**)

Χαρακτηριστικά Μάθησης ΤΝΔ

- **Υποπροσαρμογή** ή ατελής μάθηση (underfitting)
- **Υπερπροσαρμογή** (overfitting)



Χαρακτηριστικά Μάθησης ΤΝΔ

- **Υποπροσαρμογή** (underfitting) ή ατελής μάθηση
  - τα δεδομένα εκπαίδευσης **δεν επαρκούν** ή **δεν είναι αντιπροσωπευτικά** της σχέσης που θέλουμε να μάθει το ΤΝΔ ή το ΤΝΔ δεν είναι κατάλληλο (π.χ. σε μέγεθος ή δομή)
- **Υπερπροσαρμογή** (overfitting)
  - πολύ υψηλό ποσοστό επιτυχίας στα δεδομένα με τα οποία εκπαιδεύτηκε το ΤΝΔ, **αλλά χαμηλής ποιότητας πρόβλεψη για άλλα άγνωστα input** από το ίδιο πρόβλημα – φαινόμενα απομνημόνευσης των δεδομένων εκπαίδευσης!
    - το ισοδύναμο της "παπαγαλίας" στους ανθρώπους – γνώση π.χ. του πώς λύνονται συγκεκριμένες ασκήσεις, αλλά αδυναμία επίλυσης άγνωστων αλλά παρόμοιων ασκήσεων.
    - Γι’ αυτό η εκπαίδευση σταματά σε μικρό σφάλμα (πχ 5%) και όχι σε πολύ μικρό ή μηδενικό!

Δεδομένα Μάθησης ΤΝΔ

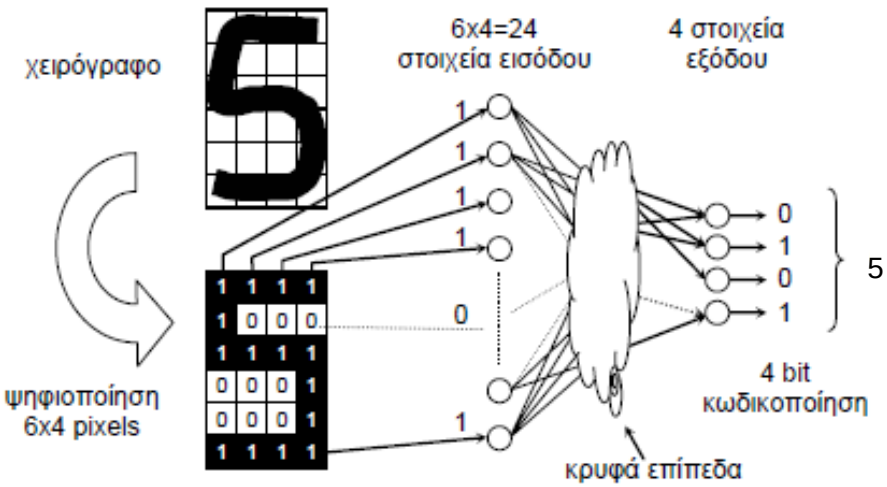
- Χρήση σε **κύκλους μάθησης/εκπαίδευσης** που ονομάζονται **εποχές** (**epochs**)
  - **μάθηση δέσμης** (batch learning)
  - **επαυξητική μάθηση** (incremental learning)
  - συνδυασμός των δύο παραπάνω μεθόδων
- Η εκπαίδευση τερματίζεται όταν το κριτήριο ελέγχου της ποιότητας του δικτύου φτάσει σε κάποια επιθυμητή τιμή.
- **Κριτήρια Ελέγχου Ποιότητας**
  - **μέσο σφάλμα** του συνόλου εκπαίδευσης
  - **μεταβολή** του **μέσου σφάλματος** του συνόλου εκπαίδευσης
- Κανονικοποίηση δεδομένων εκπαίδευσης και ελέγχου
  - (τα δεύτερα, με βάση τις παραμέτρους κανονικοποίησης των πρώτων).

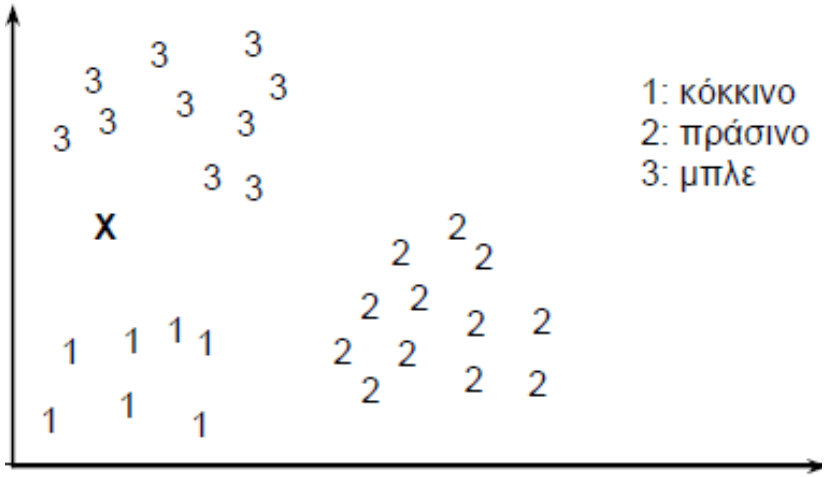
Βασικές Ιδιότητες των ΤΝΔ

- Η ικανότητα τους να **μαθαίνουν μέσω παραδειγμάτων**
  - **learn by example**
- Η δυνατότητα θεώρησής τους ως **κατανεμημένη μνήμη** (**distributed memory**) και ως **μνήμη συσχέτισης** (**associative memory**).
- Η μεγάλη τους **ανοχή σε σφάλματα**
  - **fault-tolerance**
- Η εξαιρετική ικανότητά τους για **αναγνώριση προτύπων**
  - **pattern recognition**
  - π.χ. οπτική αναγνώριση χαρακτήρων ή χειρόγραφου

ΤΝΔ Πρόσθιας Τροφοδότησης (**feedforward**)

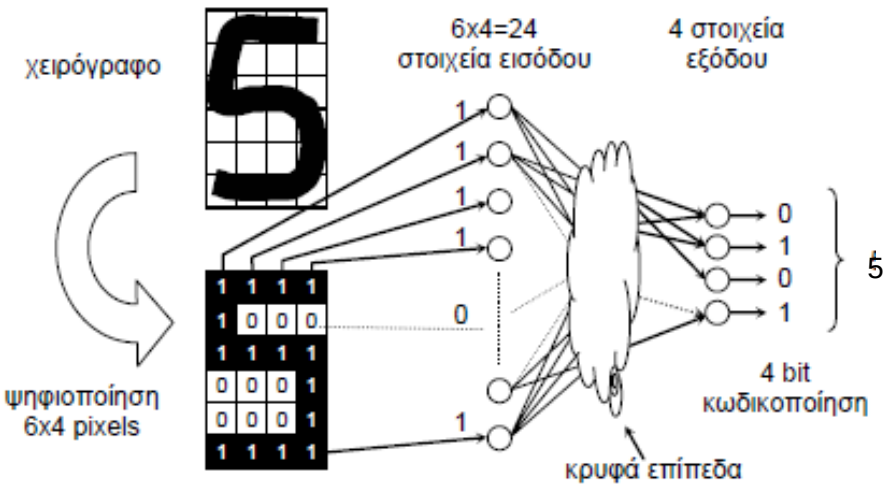
- Επίπεδα: (i) εισόδου, (ii) εξόδου, (iii) κανένα, ένα ή περισσότερα κρυφά επίπεδα
- **Είδος μάθησης**: μάθηση με επίβλεψη.
- **Τοπολογία** του δικτύου
  - Δεν υπάρχει κανόνας για τον προσδιορισμό κρυφών επιπέδων, νευρώνων ανά επίπεδο, συνδεσμολογίας
  - Τα δεδομένα εισόδου-εξόδου βοηθούν στην **εκτίμηση του αριθμού νευρώνων στα επίπεδα εισόδου** και **εξόδου**.

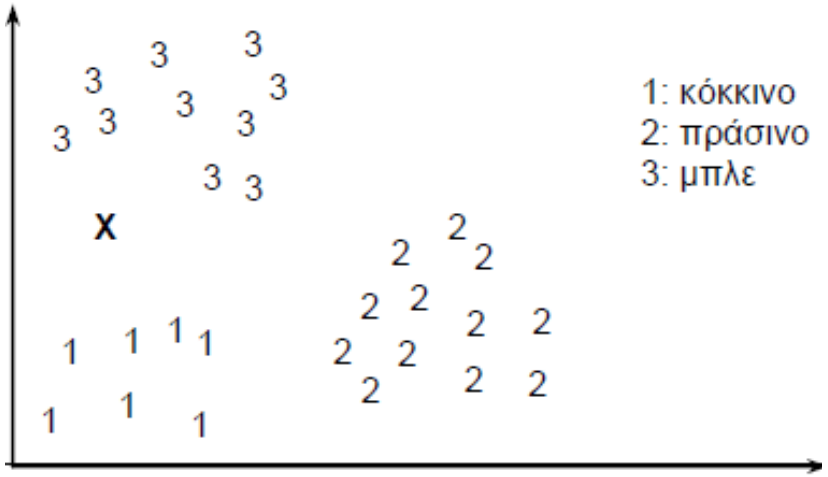




ΤΝΔ Πρόσθιας Τροφοδότησης (**feedforward**)

- Τα δεδομένα εισόδου-εξόδου βοηθούν στην εκτίμηση του αριθμού νευρώνων στα επίπεδα εισόδου και εξόδου.
- Π.χ. **για αναγνώριση** των **ψηφίων 0-9** συνίστανται:
  - 10 νευρώνες εξόδου (ένας για κάθε ψηφίο – π.χ. 0=0000000000, 1=0100000000, 2=0010000000, 3=0001000000, κ.ο.κ.), **ή**
  - **4 νευρώνες** (βλ. περίπτωση σχήματος), εφόσον όμως οι αριθμοί κωδικοποιηθούν σε αναπαράσταση 4-bits (0=0000, 1=00001, 2=0010, 3=0011, κ.ο.κ.)
- Π.χ. **για αναγνώριση 3 κατηγοριών** χωρίς σειρά (όπως **χρώματα**) συνίστανται 3 νευρώνες εξόδου, γιατί: αν 1=κόκκινο και 3=μπλέ ένα ΤΝΔ με έναν νευρώνα στην έξοδο θα έδινε απάντηση για την άγνωστη περίπτωση X (βλ. σχήμα) ότι ανήκει στην κατηγορία 2 (μέσος όρος 1 και 3), δηλαδή κατηγορία "πράσινο", που στο παράδειγμα του διαγράμματος είναι λάθος, καθώς η θέση του x υποδεικνύει ότι είναι ισοπίθανα κόκκινο ή μπλε.





Κρυφά Επίπεδα

- Ο **αριθμός των νευρώνων στα κρυφά επίπεδα σχετίζεται με πολύπλοκο τρόπο με**:
  - τον **αριθμό των νευρώνων** στα επίπεδα εισόδου και εξόδου,
  - τον **αριθμό των διανυσμάτων εκπαίδευσης** και την ύπαρξη ή όχι **θορύβου** σε αυτά,
  - την **πολυπλοκότητα** της **συνάρτησης** ή της κατηγοριοποίησης που πρέπει να μάθει το ΤΝΔ
  - τις **συναρτήσεις ενεργοποίησης** που χρησιμοποιούνται,
  - τον **αλγόριθμο εκπαίδευσης**, κ.α.
- **Εμπειρικός κανόνας για προβλήματα κατηγοριοποίησης**:
  - **αριθμός νευρώνων** στα κρυφά επίπεδα **< αριθμό διανυσμάτων εκπαίδευσης**
  - **αιτία**: για να αποφευχθεί η απομνημόνευση
    - πολλοί νευρώνες -> πολλά βάρη w -> πολλοί βαθμοί ελευθερίας του συστήματος -> ευκολία στο να απομνημονεύσει αντί να μάθει, δηλ. να συσχετίσει μέρος του δικτύου με δεδομένο διάνυσμα εκπαίδευσης!
- Συνήθως, κάθε νευρώνας συνδέεται με όλους τους νευρώνες του επόμενου επιπέδου.
- Εν γένει απαιτούνται **αρκετές δοκιμές και πειραματισμοί**.

Perceptron 1/2

- Η πιο απλή τοπολογία δικτύου με απλή τροφοδότηση.
  - **1** νευρώνας (!), με **βηματική** συνάρτηση ενεργοποίησης, μάθηση **με επίβλεψη**
  - Πώς συντελείται η μάθηση: **μέσω αναπροσαρμογής των τιμών των βαρών**.
    - Αρχικά τα βάρη έχουν **τυχαίες τιμές**
    - Τις αναπροσαρμόζουν επιδιώκοντας για δεδομένο input να υπολογίσουν το (γνωστό) επιθυμητό output – οδηγός της αναπροσαρμογής είναι το σφάλμα μεταξύ της τιμής που υπολογίζουν και της τιμής που θέλουμε να υπολογίσουν (με άλλα λόγια μαθαίνουν από τα λάθη τους!)
    - Τα δεδομένα (input-output) που χρησιμοποιούνται ονομάζονται **δεδομένα εκπαίδευσης** (training data) και προέρχονται **από μετρήσεις/καταγραφή** αυτού που θέλουμε να μάθουν
      - π.χ. μαθαίνουν να κάνουν διάγνωση καρκίνου εκπαιδευόμενα με δεδομένα εξετάσεων υγειών ατόμων και ατόμων που αποδεδειγμένα πάσχουν/έπασχαν από την ασθένεια
    - στις γενικές του αρχές ο μηχανισμός ισχύει και για πιο πολύπλοκα νευρωνικά δίκτυα
  - Μπορούμε να συνδυάσουμε perceptrons για να φτιάξουμε πιο πολύπλοκα ΤΝΔ

Perceptron 2/2

- Αλγόριθμος αναπροσαρμογής (μεταβολής) των βαρών
  - Επανάληψη έως ότου ικανοποιηθεί η **συνθήκη τερματισμού** της εκπαίδευσης:
  - **Για κάθε ζευγάρι εισόδου x** και επιθυμητής **εξόδου t** από το σύνολο δεδομένων εκπαίδευσης
  1. **Υπολόγισε** την έξοδο y
  2. Εάν y=t, τότε δε γίνεται καμία μεταβολή στα βάρη
  3. Εάν y≠t, τότε μετέβαλε τα βάρη κατά την ποσότητα $\Delta w = d \cdot (t-y) \cdot x$, έτσι ώστε το y να πλησιάσει το t.
- d: σταθερά - ρυθμός μάθησης (learning rate) – συνήθως έχει μικρές τιμές.

Μνήμες Συσχέτισης (Associative Memories)

- Συστήματα μνήμης που ορίζουν απεικονίσεις μεταξύ δύο αναπαραστάσεων X και Y, έτσι ώστε όταν δοθεί η μία, να μπορεί να ανακληθεί η άλλη.

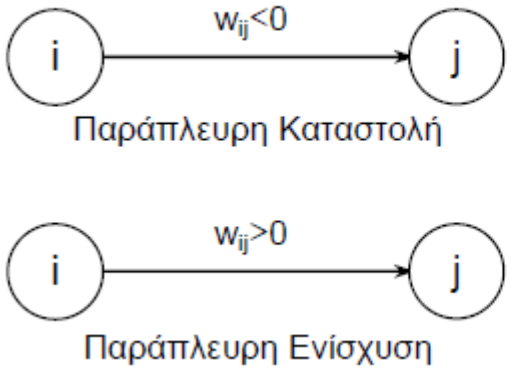
- Ανάλογα με τις **διαφορές μεταξύ εισόδου και εξόδου** διακρίνουμε:
  - **αυτοσυσχετιζόμενες** μνήμες (auto-associative memories)
  - **ετεροσυσχετιζόμενες** μνήμες (hetero-associative memories)

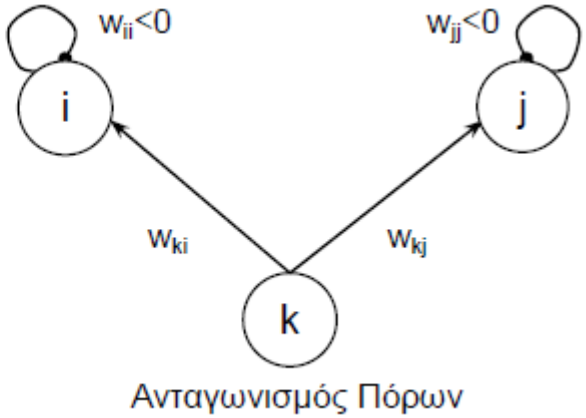
- Ανάλογα με το αν η **έξοδός** τους μπορεί να είναι **προϊόν παρεμβολής**, διακρίνουμε:
  - **με** δυνατότητα παρεμβολής (interpolative associative memories)
  - **προσαυξητική μνήμη συσχέτισης** (accretive associative memory)

- Τύποι ΤΝΔ που συνιστούν μνήμες συσχέτισης:
  - Γραμμικοί Συσχετιστές, **Δίκτυα Hopfield**, **Μνήμες Συσχέτισης Διπλής Κατεύθυνσης**

ΤΝΔ με Ανταγωνισμό

- **Βασική Ιδέα**: οι νευρώνες πρέπει να είναι σε θέση να επηρεάσουν θετικά, ουδέτερα ή ακόμη και αρνητικά τους υπόλοιπους νευρώνες του δικτύου.
  - ποιος νευρώνας θα ανταποκριθεί περισσότερο;
  - απλούστερη περίπτωση: μόνο ο νευρώνας με τη μεγαλύτερη έξοδο (νικητής) παράγει τελικά αποτέλεσμα (winner-takes-all - WTA).
- Μοντελοποίηση Ανταγωνισμού
  - **Παράπλευρη καταστολή** ή **ενίσχυση** (lateral inhibition ή excitation)
  - **Ανταγωνισμός πόρων** (resource competition)
    - βιολογικά αποδεκτό - μοντελοποιεί και το βιολογικό φαινόμενο της εξασθένισης (self decay).





Εφαρμογές Νευρωνικών Δικτύων 1/4

- Δημοφιλή σε προβλήματα που περιέχουν μη-προβλέψιμες λειτουργίες και τα οποία δεν είναι πλήρως κατανοητά ώστε να δοκιμαστούν μαθηματικά μοντέλα.
- Κατηγοριοποίηση
  - **Ιατρική**
    - Κατηγοριοποίηση ιατρικών εικόνων που προέρχονται από εξετάσεις υπέρηχων, ηλεκτροκαρδιογραφήματα, τεστ Παπανικολάου, κτλ.
    - Τα ΤΝΔ καλούνται να κάνουν μια πρώτη διάγνωση, επιταχύνοντας σημαντικά τη χρονοβόρα διαδικασία ελέγχου των δεδομένων ιατρικών εξετάσεων από τους ιατρούς.
    - Οι περιπτώσεις που κρίνονται ως ύποπτες, εξετάζονται στη συνέχεια από ιατρούς.

Εφαρμογές Νευρωνικών Δικτύων 2/4

- Κατηγοριοποίηση
  - **Τομέας άμυνας**: κατηγοριοποίηση εικόνων προερχόμενων από radar, sonar, κτλ.
  - **Γεωργία**: έλεγχος καλλιεργειών σε συνδυασμό με δορυφορικές εικόνες
  - **Οικονομία/επιχειρήσεις**: κατηγοριοποίηση πελατών βάσει αγοραστικών τους συνήθειων
- Αναγνώριση
  - **Τράπεζες**: γνησιότητα υπογραφής και χαρτονομισμάτων
  - **Πληροφορική και Τηλεπικοινωνίες**: αναγνώριση ήχου, εικόνας και γραπτού κειμένου (OCR), Google Translate (http://www.nooz.gr/article/to-google-translate-dimioirgise-diki-tou-glossa)

Εφαρμογές Νευρωνικών Δικτύων 3/4

- Αποτίμηση
  - **Τομέας άμυνας**: παρακολούθηση στόχων.
  - **Ασφάλεια**: εντοπισμός κίνησης (motion detection), ταύτιση δακτυλικών αποτυπωμάτων, ανάλυση εικόνας σε συστήματα επιτήρησης.
  - **Μηχανολογία**: παρακολούθηση, επιθεώρηση και έλεγχος προϊόντων.
- Πρόβλεψη
  - **Οικονομία/επιχειρήσεις**: πρόβλεψη ισοτιμίας νομισμάτων και τιμών μετοχών (συνήθως βραχυπρόθεσμη), πρόβλεψη πωλήσεων, κτλ.

Εφαρμογές Νευρωνικών Δικτύων 4/4

- Πρόβλεψη
  - **Γεωργία**: πρόβλεψη παραγωγής, κυρίως με χρήση δορυφορικών εικόνων (π.χ. σιτηρά που πρόκειται να παράγουν πολύ έχουν διαφορετική απεικόνιση από αυτά που δεν θα αποδώσουν – η φωτογράφιση γίνεται συνήθως στο υπέρυθρο φάσμα)
  - **Μετεωρολογία**: μοντέλα πρόβλεψης καιρού

Σύνοψη Νευρωνικών Δικτύων

- Αν και ένα εκπαιδευμένο ΤΝΔ μπορεί να αναγνωρίσει δεδομένα τα οποία δεν έχει δει ποτέ του, αυτό δεν συμβαίνει στην περίπτωση που τα δεδομένα δεν ανήκουν στο ίδιο φαινόμενο/πρόβλημα με δεδομένα του οποίου έχει εκπαιδευτεί !!!

- **Δεν υπάρχουν ΤΝΔ γενικού σκοπού** τα οποία να μπορούν να αντιμετωπίζουν διάφορα ετερογενή προβλήματα.

- Υπάρχουν **πολλά είδη** νευρωνικών δικτύων (π.χ. Kohonen) με κάθε είδος να έχει ιδιαίτερες ικανότητες και χαρακτηριστικά (π.χ. λειτουργία χωρίς εκπαίδευση).

Ερωτήσεις - Απορίες